CHAPTER 5
INTRODUCING COSMOLOGY – THE SCIENCE OF THE UNIVERSE

5.1 Introduction

Cosmology is the branch of science concerned with the study of the Universe as a whole. It involves questions such as: 'What is the composition of the Universe?'; 'What is its structure?'; 'How did it originate?'; 'How is it evolving?' and 'What is its ultimate fate?' These are obviously very challenging questions. They have been subjects of religious and philosophical speculation for thousands of years, but the development of scientific cosmology has brought them into the mainstream of astronomical debate over the past hundred years or so, and there is now real hope that their answers are coming into view.

Cosmology is a huge subject, and much of it concerns vast scales of time and distance. However, as you will see, cosmology has to concern itself with the physics of the very small (i.e. the physics of subatomic particles) as well as the physics of the very large. It is this combination of the very small and the very large – the microscopic and the macroscopic – that gives modern cosmology its distinctive flavour. It is also this combination that enables cosmology to cast new light on the nature of matter, and this gives it a vital role to play in answering one of the hardest questions in contemporary astronomy – 'Where did the galaxies come from?'

This chapter provides a broad introduction to scientific cosmology. Section 5.2 sets the scene by drawing together a number of facts about the Universe, most of which you have met in earlier chapters. Section 5.3 is concerned with 'modelling' the Universe: the process of formulating simplified descriptions of the Universe, usually expressed in mathematical form, that are consistent with modern physics, particularly with Einstein's theory of gravity – the general theory of relativity. The process of modelling the Universe involves a number of important cosmological parameters, such as the age of the Universe, the total density of matter in the Universe, and the Hubble constant, H_0. Section 5.4 concludes this introduction to cosmology by highlighting these cosmological parameters and considering the relationships between them that are predicted by the most popular models of the Universe. Later chapters build on this introduction by considering, in turn, the nature of the early Universe and the big bang (Chapter 6), the challenges and results of measuring the key cosmological parameters (Chapter 7), and the many important questions that are still unanswered at the current stage in the development of cosmology (Chapter 8).

5.2 The nature of the Universe

This section briefly introduces some of the main facts about the Universe, as revealed by astronomical observations. Many of these facts should already be familiar to you from earlier chapters, but some will be new. In neither case will much be said here about how the information was obtained. The main aim is simply to catalogue the basic facts that must be explained by any theoretical account of the origin and evolution of the Universe. Where necessary, greater detail is given later.

Before giving the observational 'facts' there is one important point that deserves special emphasis. All of our observations of the Universe are carried out from points on or near the Earth. In interpreting astronomical observations we have learned from experience *not* to assume that the Earth is in any particularly privileged position. We are not at the centre of the Solar System, nor are we at the centre of the Milky Way. It seems reasonable, therefore, to suppose that we are not at the centre of the Universe either. (This is absolutely opposed to the ancient pre-scientific view that placed us at, or close to, the centre of the Universe.) The assumption that we do not occupy a privileged position in the Universe is usually referred to as the **Copernican principle**, and is often invoked in interpreting observational data. If, for example, we find that distant galaxies are heading away from us in all directions (as we do), the Copernican principle tells us that the observations do not mean that we have the privilege of being at the fixed centre of an expanding Universe, but rather that the nature of cosmic expansion is such that the recession of distant galaxies is what would be observed from *any* typical point in the Universe.

The term 'Copernican principle' recalls the work of Nicolaus Copernicus (1473–1543), who proposed a Sun-centred model of the Solar System at a time when the prevailing view favoured an Earth-centred model.

5.2.1 The matter in the Universe

One of the most obvious facts about the Universe is that it contains matter. We humans are made of matter, as are the planets, stars, nebulae and galaxies that we observe. All of these visible objects are basically composed of oppositely charged *electrons* and *nuclei*. In some cases, such as ourselves and most of the Earth, the electrons and nuclei are combined together to form electrically neutral *atoms*, but the major part of the visible matter – most of that in stars for example – takes the form of a *plasma* in which the electrons and nuclei are separate and distinct, although the plasma as a whole remains electrically neutral.

Wherever large bodies of visible matter are found, whether they are composed of atoms or plasma, it is always the case that their mass is mainly accounted for by the *protons* and *neutrons* that make up nuclei, since these particles are far more massive than the electrons that accompany them.

■ The masses of the electron, proton and neutron are: $m_e = 9.109 \times 10^{-31}$ kg, $m_p = 1.673 \times 10^{-27}$ kg and $m_n = 1.675 \times 10^{-27}$ kg. Use these values to evaluate the following ratios: m_p/m_n, m_p/m_e and m_n/m_e. Roughly what fraction of the mass of a helium atom is attributable to the protons and neutrons contained in its nucleus?

❏ The required ratios are $m_p/m_n = 0.9988$, $m_p/m_e = 1837$ and $m_n/m_e = 1839$. Taking the view that the helium nucleus has the combined mass of two protons and two neutrons, while the helium atom has the combined mass of the nucleus and two electrons, we should expect the fraction of atomic mass attributable to protons and neutrons to be

$$(2m_p + 2m_n)/(2m_p + 2m_n + 2m_e) = 6696/6698 = 0.9997$$

(This is only an approximate value since we have ignored the effects of *binding energy*, which causes the mass of the nucleus to be slightly less than that of two protons and two neutrons.)

As you saw in Chapter 1, both the proton and the neutron (but not the electron) belong to a family of elementary particles called *baryons*, and the term *baryonic matter* is used to refer to all forms of matter in which the mass is mainly attributable to baryons. All the familiar atomic and molecular gases, liquids and solids, whatever their chemical composition, and all the plasmas found in stars, nebulae and galaxies, are therefore examples of baryonic matter.

Analyses of the chemical compositions of stars, nebulae and galaxies indicate that the most common form of baryonic matter is actually hydrogen plasma, and the second most common form is helium plasma. Although there are other significant forms of baryonic matter, we can, very crudely, say that the Universe has about 75% of its baryonic mass in the form of hydrogen nuclei and about 25% in the form of helium nuclei. We refine this crude recipe in later chapters, but it's worth remembering it as a first approximation to the true chemical composition of the Universe. Explaining the relative abundance of hydrogen and helium is a major challenge for any theory of the Universe, and is addressed in Chapter 6.

- Accepting that 75% of the baryonic mass of the Universe is due to hydrogen and 25% is due to helium, what are the relative numbers of hydrogen and helium nuclei in the Universe? In particular, how many hydrogen nuclei would you expect to find for each helium nucleus?

- Representing the masses of the hydrogen and helium nuclei by m_{H} and m_{He}, respectively, and using the symbols n_{H} and n_{He} to represent the number densities of hydrogen and helium nuclei, the question implies that

$$\frac{m_{\mathrm{H}}\,n_{\mathrm{H}}}{m_{\mathrm{He}}\,n_{\mathrm{He}}} = \frac{75}{25} = 3$$

Making the approximation that $m_{\mathrm{He}} = 4m_{\mathrm{H}}$ it follows that

$$\frac{n_{\mathrm{H}}}{4n_{\mathrm{He}}} = 3\text{, or equivalently, }\frac{n_{\mathrm{H}}}{n_{\mathrm{He}}} = 12$$

So, there should be 12 hydrogen nuclei for each helium nucleus.

It's clear that the Universe contains a great deal of baryonic matter, since it contains a lot of luminous stars, nebulae and galaxies. But earlier chapters have emphasized that many independent observations also indicate the presence of a great deal of non-luminous dark matter that has so far been detected only through its gravitational influence. Some scientists have suggested that dark matter may not actually exist at all, and that it may be our understanding of gravity that is at fault. However, the majority view at the time of writing is that dark matter does exist, and that it is far more common than baryonic matter. As stated in Chapter 1, it is expected that some of the dark matter is non-luminous *baryonic dark matter*, but there are good reasons to believe that most of it is non-baryonic, and that this *non-baryonic dark matter* is, at least in terms of density, the dominant form of matter in the Universe.

Recent cosmological observations indicate that the total density of non-baryonic dark matter is between five and six times greater than that of baryonic matter (see Figure 5.1). These measurements are discussed in Chapter 7, while the reasons for expecting only a limited amount of baryonic matter are covered in Chapter 6.

Figure 5.1 Galaxies represent large concentrations of matter. The visible parts of galaxies are certainly composed of baryonic matter, but most of the matter in galaxies is thought to be dark matter and the majority of that is expected to be non-baryonic. (Hubble Heritage Team (AURA/STScI/NASA))

QUESTION 5.1

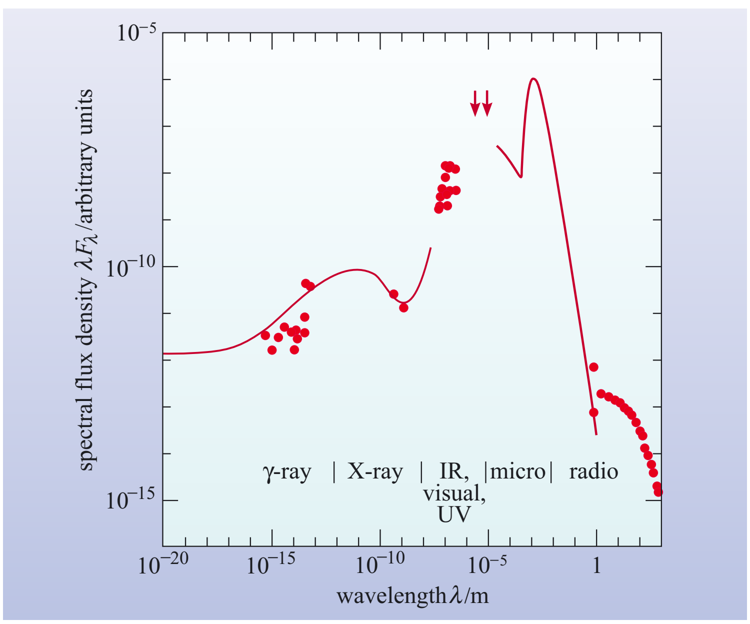
For every 10 neutrons in the Universe, how many protons would you expect to find? Explain the assumptions you have made in arriving at your answer.

5.2.2 The radiation in the Universe

Another important fact about the Universe is that it is essentially full of *electromagnetic radiation*. All that we know of distant galaxies and clusters has been learnt by observing electromagnetic radiation (mainly radio waves, infrared radiation, visible light, X-rays and γ-rays) that has originated in those galaxies or clusters and then travelled through space until detected here on Earth. Of course, this direct observation only shows that there is a lot of radiation travelling towards the Earth, but the Copernican principle tells us that the Earth does not occupy any special place in the Universe, hence the belief that all parts of the Universe receive radiation from their cosmic surroundings, and the statement that the Universe is essentially full of radiation. This is supported by observations of the effects of the radiation on distant bodies.

One of the characteristic properties of electromagnetic radiation is its wavelength, λ, and an important feature of any observed radiation is its *spectral energy distribution* (defined in Chapter 3), which provides a description the amount of energy delivered per second and per unit area in any narrow range of wavelengths belonging to the radiation concerned. The radiation that reaches the Earth from space is dominated by the radiation from the Sun, and this is dominated by the visible light that our eyes have evolved to observe. However, this dominant role of visible light in our neighbourhood is a result of our close proximity to the Sun. Observations indicate that, in the Universe as a whole, the spectral energy distribution of radiation is actually dominated by microwave radiation, which occupies a range of wavelengths around 1 mm, between radio waves and infrared radiation.

Figure 5.2 The spectral energy distribution of background radiation. (Adapted from Sparke and Gallagher, 2000; based on work by T. Ressell and D. Scott)



The predominance of microwave radiation is indicated in Figure 5.2, which shows the spectral energy distribution of so-called ‘background radiation’ at various wavelengths. This background radiation does not come from any identified source (such as the Sun or the Moon), and is in part thought to be truly ‘cosmic’ in the sense that similar distributions would be seen from anywhere in the Universe at the present time. The microwave contribution to the background radiation is usually referred to as the **cosmic microwave background (CMB)** and, at least in terms of the energy it carries, represents the dominant form of radiation in the Universe.

The discovery of the CMB, in 1965, was one of the most important events in the development of scientific cosmology and is described in Chapter 6. As you will see, it did a great deal to establish the ‘big bang theory’ as the best supported theory of cosmic evolution, and it continues to be of the utmost importance since detailed observations of the CMB now provide precise information about the nature, composition and evolution of the Universe.

5.2.3 The uniformity of the Universe

It is an obvious fact that what is normally above your head (probably some air, a ceiling, a roof, a lot more air and then outer space) is different from what is below your feet (probably a floor, some building foundations and the whole of the Earth). This provides clear evidence that, locally at least, the Universe is *not* uniform. However, rather than concentrate on your immediate surroundings, think instead about a sufficiently large region of the Universe that an astronomer or cosmologist might regard it as a 'typical' or 'representative' sample. In practice this means considering a region that is large enough to contain a few superclusters of galaxies, along with the voids that typically separate such superclusters. Such a region might well have a diameter of several hundred megaparsecs, and might represent, say, a millionth of the total volume of the observable part of the Universe. Provided the temptation to think about regions smaller than this is resisted, the view of modern cosmologists is that any part of the Universe is pretty much like any other at the present time. That is to say that, if we consider any two regions of the present Universe, each sufficiently large to be representative, then those regions will have the same average density, pressure, temperature, etc. In this sense, cosmologists say, provided we consider sufficiently large size scales, the Universe is uniform.

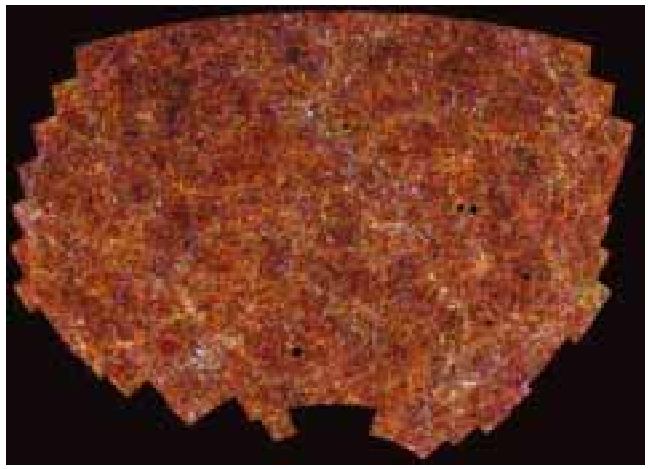


Figure 5.3 The observed distribution of galaxies across about 4300 square degrees of sky around the South Galactic Pole. (Steve Maddox, APM Galaxy Survey)

Belief in the large-scale uniformity of the Universe has always played an important part in scientific cosmology. In the early days of the subject this belief was based on assumption and the absence of any contradictory evidence, but in recent years it has come to rest more and more on positive observational support. One thing that observation certainly makes clear is that, on the large scale, the Universe is pretty much the same in all directions. This is fairly clear just from the large-scale distribution of galaxies, which can be seen to be reasonably even in all directions that are not obscured by parts of the Milky Way (see Figure 5.3). However, even better evidence is provided by the cosmic microwave background (see Figure 5.4), which is observed to come with equal intensity (to about a few parts in 10^5) from all regions of the sky, once allowance has been made for the 'local' effects of the Earth's motion.

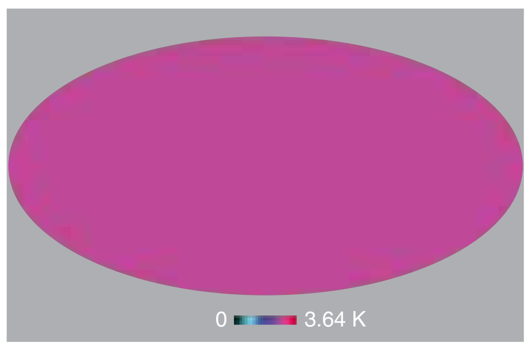


Figure 5.4 The observed intensity of the cosmic microwave background radiation across the whole sky. These data come from space-based measurements by the COBE satellite (Cosmic Background Explorer) and have been corrected to compensate for the motion of the detector. They are based on measurements made with an angular resolution of a few degrees and are sensitive to intensity variations of about one part in a thousand. (Douglas Scott)

By combining the evidence that the large-scale Universe appears the same in all *directions* when observed from our location, with the Copernican principle that we are not in any 'privileged' location, it follows that the large-scale Universe should appear to be the same in all directions from *every* location. This in turn implies that it should be the same everywhere, that is to say it should be uniform. This combination of observation and assumption is quite convincing in itself, but in recent years even more support has been provided by the increasingly ambitious surveys of galaxy redshifts that were described in Chapter 4. These really do provide evidence that galaxies are distributed uniformly on the large scale. There are signs of clustering and superclustering on scales of tens or hundreds of

megaparsecs, but there is no sign of any larger scale structure. It is *not* the case, for example, that the galaxies on one side of the Earth are significantly more clustered than those on the other side. The evidence for large-scale uniformity was discussed in detail in Chapter 4 and some results from the recent 2dF survey were shown in Figure 4.26. Other results from this survey are shown in Figure 5.5.

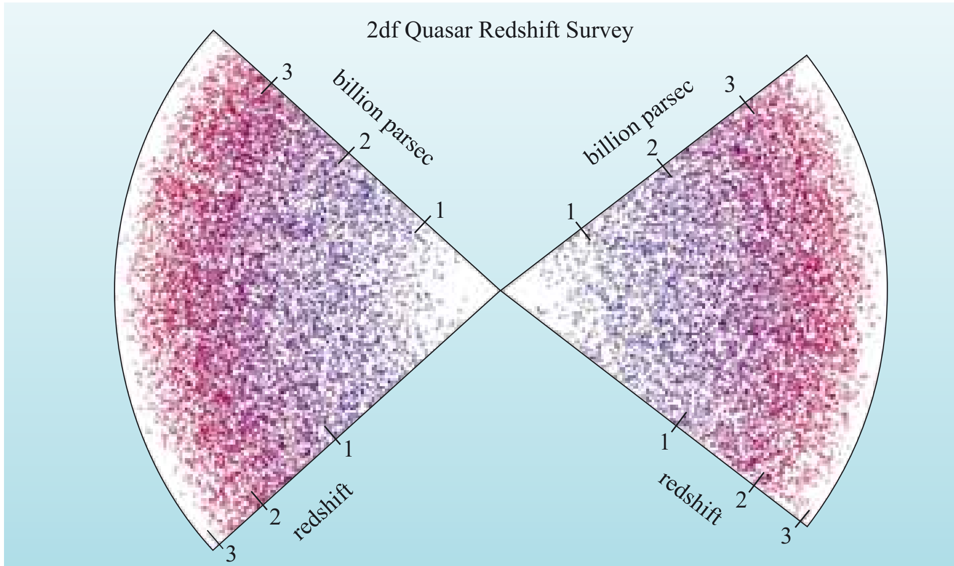


Figure 5.5 The distribution of quasars in two thin, wedge-shaped slices of the Universe. The quasars are observed out to such large distances that evolutionary effects allow changes in the number of quasars per unit volume to be observed as distance from the Earth increases. However, at any given distance the data give strong support to the claim that the large-scale distribution of quasars is the same in all directions. (Robert Smith, 2dF Quasar Redshift Survey)

QUESTION 5.2

In assuming that we can use the Copernican principle to interpret our observations of the CMB we are assuming that the CMB is a truly cosmic phenomenon, rather than a purely local one such as, say, sunlight. Describe a piece of observational evidence that makes it plausible to suppose that the CMB is a cosmic phenomenon, whereas sunlight is only a local astronomical one.

5.2.4 The expansion of the Universe

The nearest galaxies to the Milky Way are mainly dwarf galaxies that appear to be 'satellites' orbiting our own Galaxy. The closest large spiral galaxy – the Andromeda Galaxy, M31 – is actually heading towards us. But looking deeper into space it is found that all distant galaxies have a component of velocity which is directed away from the Earth, as revealed by the redshifts seen in their spectra. This finding, by Vesto Slipher (1875–1969), paved the way for the discovery of Hubble's law which was described in Chapter 2. As explained in Chapter 2, Hubble's law applies to galaxies with redshifts up to about 0.2, and implies a direct proportionality between redshift and distance that can be written as

$$z = \frac{H_0}{c} d \quad (\text{for } z \lesssim 0.2) \tag{5.1}$$

where H_0 is Hubble's constant and c is the speed of light in a vacuum. The redshift z can be related to the observed and emitted wavelengths, λ_{obs} and λ_{em}, of some identified spectral line by the equation (this is a repeat of Equation 2.11)

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} \tag{5.2}$$

The two parts of Figure 5.6 provide a visual reminder of the meaning of redshift, as well as an indication of some of the observational data that support Hubble's law.

For reasons that will become clear later, it would be a mistake to interpret the redshift seen in the spectra of distant galaxies as a simple Doppler effect. Nonetheless, it is true that the observed redshifts do indicate that all distant galaxies are receding even though the Doppler formula can't generally be used to determine the speed of that recession. As mentioned earlier, such an observed recession is not thought to prove that the Earth is at the centre of an expanding cloud of galaxies, but rather that the whole Universe is in a state of expansion, with every galaxy, on average, moving away from every other *distant* galaxy. This overall expansion, described by Hubble's law, is sometimes called the **Hubble flow**. Galaxies provide imperfect tracers of the Hubble flow, since they interact gravitationally and may therefore be disturbed by the presence of nearby galaxies or other local effects. These local disturbances manifest themselves as movements relative to the Hubble flow, called *peculiar motions*, and are believed to be the main reason why a plot of redshift against distance, even for distant galaxies, is not the perfect straight line

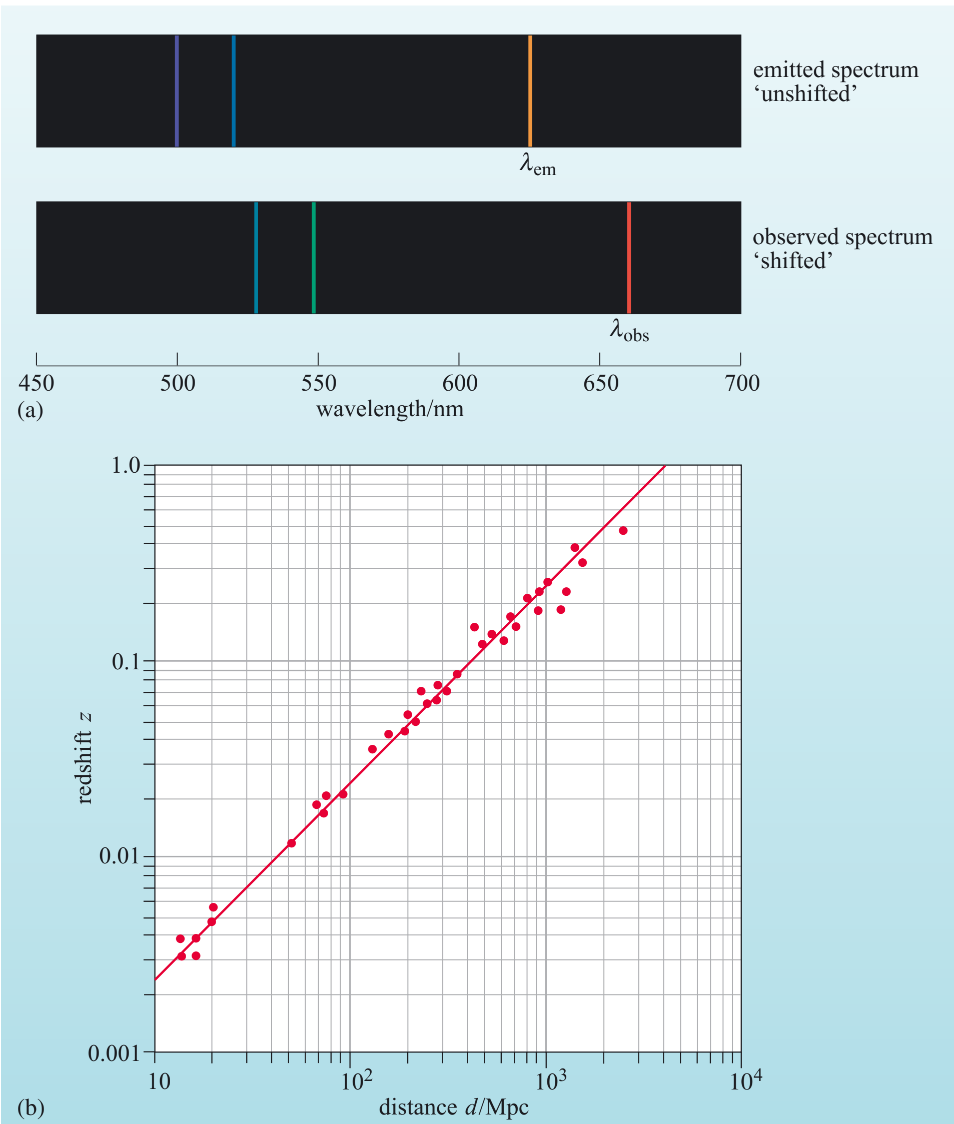


Figure 5.6 (a) The redshift z of a spectral line observed at wavelength λ_{obs} is a fractional measure of the line's displacement from the wavelength λ_{em} at which it was emitted by its source. (b) Observational evidence in support of Hubble's law.

that Hubble's law implies. The expansion of the Universe, like the uniformity of the Universe, is a large-scale phenomenon, and local departures such as the approach of M31 are only to be expected.

It's worth noting that if the uniformity of the Universe is to be preserved over time then the expansion of the Universe must also be uniform. At any time, the Universe should be expanding equally in all directions when viewed from any typical point. As you will see later, in an expansion described by Hubble's law, the current rate of expansion is measured by the Hubble constant, H_0. The uniformity of the Universe therefore implies that the Hubble constant should be a 'universal constant' with a value that is independent of where it is measured, though not necessarily of when it is measured. We have a lot more to say about this important constant later in our discussion.

■ Summarize the four main facts about the present state of the Universe that have been discussed in this section, giving detailed clarification where appropriate.

❑ The Universe contains matter. The matter is mainly non-baryonic dark matter, but about 1/5th or 1/6th is baryonic matter, mainly hydrogen (~75%) and helium (~25%).

The Universe contains radiation. The radiation is mainly cosmic microwave background radiation.

The Universe is uniform. All regions that are sufficiently large to be representative currently have the same average density, wherever they are located. This is consistent with the observed distributions of matter and radiation.

The Universe is expanding. For redshifts of ~0.2 or less, the expansion is thought to be well described by Hubble's law, implying that the current rate of expansion is described by the Hubble constant. Galaxies are thought to provide somewhat imperfect tracers of this expansion since they may have local motions relative to the large-scale Hubble flow.

5.3 Modelling the Universe

Physicists generally take the view that a scientific understanding of a phenomenon has been achieved when that phenomenon can be accurately described in terms of a few concise statements, or better still a well-formulated equation. A typical example is provided by the flow of electric current I through a sample of electrically conducting material in response to an applied voltage V. This is described by a relationship known as *Ohm's law*, which is expressed by the equation $V = IR$, where R is a parameter, called the *resistance*, that characterizes the electrical properties of the sample. Ohm's law provides a **mathematical model** of the process of current flow, implying that, for any given sample, V is proportional to I, but requiring the value of R to be determined before it can supply quantitative predictions.

Cosmologists adopt a similar view regarding the understanding of the Universe. One of the central concerns of modern cosmology is the formulation and investigation of mathematical models of the Universe. These are called **cosmological models**. They usually take the form of a few *equations* that imply general relations between observable quantities, but they also involve *parameters*, such as the Hubble constant, that must be determined by observation before the model can be used to provide detailed quantitative predictions.

ALBERT EINSTEIN (1879–1955)

Albert Einstein (Figure 5.7) was born in Ulm, Germany in 1879. From 1896 to 1901 he lived in Zurich, Switzerland, where he was a student at the Federal Institute of Technology (ETH). In 1905 he was working in the patent office in Bern when he completed some of the most important papers in the history of physics. These included a paper on the photoelectric effect, a paper on Brownian motion, and two papers on the special theory of relativity, the second of which introduced the equation $E = mc^2$. About ten years later, as Professor of Physics in Berlin, Einstein completed his general theory of relativity, a generalization of the 1905 theory that also turned out to be a theory of gravity. General relativity received its first systematic exposition in 1916, and was first applied to cosmology in 1917.

Observations made in 1919, during a total eclipse of the Sun, confirmed one of the key predictions of general relativity, that starlight passing close to the edge of the Sun should undergo a gravitational deflection of 1.74 arcsec (see Section 4.3.2). The success of this prediction was front-page news, and made Einstein an international celebrity. He was awarded the Nobel Prize for physics in 1921 (mainly for his work on the photoelectric effect), and received many other honours and awards. In 1932, shortly before the Nazis came to power, he left Germany for the United States where he took up a post at the Institute for Advanced Study in Princeton. With war approaching, in 1939, Einstein was persuaded to sign a letter to President Roosevelt, pointing out the military implications of atomic power. In his later years Einstein became a prominent commentator on world affairs, but had little direct impact on the development of science. He pursued a bold but fruitless search for a unified field theory that would unite gravitation and electromagnetism, and in 1952 he was offered the Presidency of Israel, which he declined. He died in Princeton in 1955.

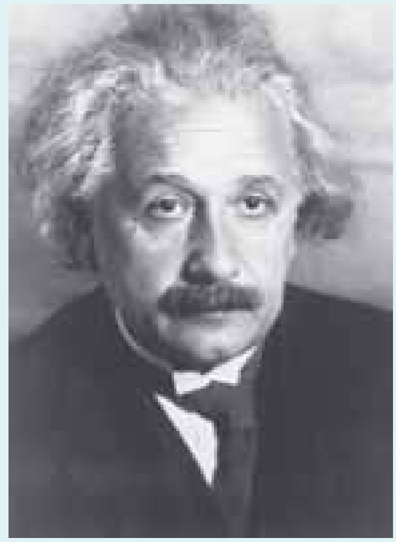


Figure 5.7 Albert Einstein. (Science Photo Library)

The aim of this section is to introduce you to some of the simplest but most important cosmological models that are currently in use, and to explore some of their implications. All of these models are based on **general relativity**, the theory of gravity published by Albert Einstein in 1916. For this reason, our discussion begins with a consideration of relativity and relativistic cosmology.

5.3.1 The relativistic Universe

The modern era of cosmology can be dated from the day in 1917 when Einstein's paper 'Cosmological Considerations of the General Theory of Relativity' was published in the Proceedings of the Prussian Academy of Science. An insight that is fundamental to this paper, and indeed to the whole of Einstein's theory of general relativity, is the crucial role that **space** and **time** must play in any attempt to model the Universe.

Before the advent of Einstein's theory of relativity, the view of most physicists was that space and time simply provided a sort of container for matter and radiation. Every particle of matter or radiation occupied some point in space at each instant of

time. Moreover, space and time were supposed to be *passive*. They provided a setting for the drama of physics, but they were not themselves players in that drama. The properties of space and time (basically geometric properties, as will be explained in the next section) were not thought to be in any way affected by the properties of the matter and radiation they contained.

Einstein changed this view radically and forever. Already, in 1905, his special theory of relativity (essentially a restricted form of the general theory that ignored gravity) had shown that the three dimensions of space and the single dimension of time should be melded together to form a unified four-dimensional entity usually referred to as **space–time**. But the general theory of 1916 went much further by showing that the geometric properties of this four-dimensional space–time were affected by the matter and radiation it contained, and that this could account for the cosmically important phenomenon of gravitation.

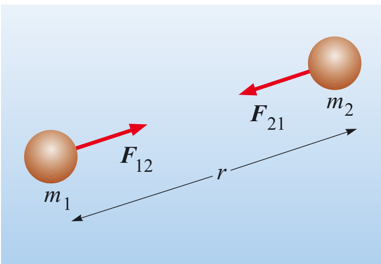


Figure 5.8 Newton's view of gravity: one body attracts another by means of a force that acts instantly across the intervening space.

According to the Newtonian theory of gravity, introduced in 1668, gravitational phenomena, such as the attraction between the Sun and the Earth, were due to a 'force' that acted instantly, between one body and another, across any intervening space (see Figure 5.8). Newton was able to describe the strength and direction of this force in terms of the masses of the bodies and the displacement (i.e. distance and direction) of one body from the other. However, he was not able to explain the origin of the gravitational force. He did not know the 'mechanism' that actually caused one body to influence another body at a remote location in space. He tried to explain his mysterious gravitational force in terms of something called 'vortex theory' that was popular with European scientists at the time, but his efforts only convinced him that this would not work, so he contented himself with *describing* the gravitational force and saying that as far as its origin was concerned 'I frame no hypothesis.' (In the Latin of his great work *Principia Mathematica*: 'hypotheses non fingo'.)

Two hundred and fifty years later Einstein was able to go much further in accounting for gravitational phenomena. According to Einstein there is no gravitational force. In Einstein's view, a body such as the Sun acts on the space–time in which it is located, giving rise to a geometrical distortion usually referred to as a **curvature** of space–time. Bodies moving in the vicinity of the Sun, such as the Earth, respond to this curvature by moving in a way that is different from the way they would have moved if the Sun had been absent and the space–time undistorted. In this way, a body such as the Sun is able to gravitationally influence the behaviour of a body such as the Earth, even though there is no 'gravitational force' acting between them. Gravitation, in Einstein's view, is a result of space–time curvature – a geometric phenomenon – and general relativity is Einstein's 'geometric' theory of gravity. This is illustrated schematically in Figure 5.9, where 'space' is represented by a two-dimensional sheet and gravity is indicated by the 'curvature' of that sheet.

Einstein was able to show that his theory of gravity agreed with all the correct predictions of Newton's theory. But general relativity went further, explaining anomalies for which Newton's theory had no account, and predicting entirely new phenomena that were outside the scope of Newtonian theory. Observations have consistently supported the novel predictions of general relativity, such as the gravitational deflection of starlight, which is why we now regard it as having superseded Newton's theory. Of course, we still use Newton's theory and speak of gravitational 'forces', but we do so because it is convenient, simple and sufficiently accurate for most purposes.

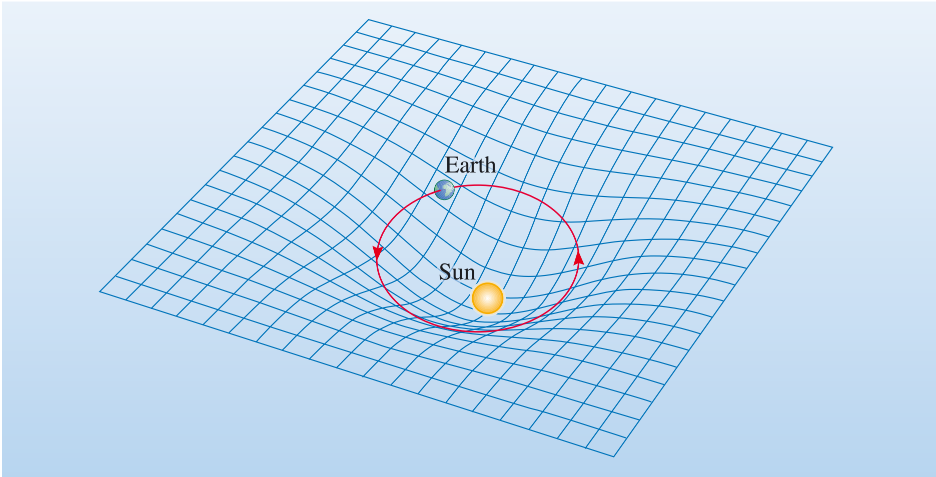


Figure 5.9 Einstein's view of gravity: a massive body (such as the Sun) significantly distorts the space–time in its vicinity. This space–time distortion (curvature) influences the motion of other bodies (such as the Earth) moving through that region of space–time, giving the impression that a 'force' is acting, although in reality there is only distorted space–time and motion in response to that space–time distortion.

What has all this to do with cosmology? Well, according to general relativity, what determines the curvature of space–time in any region of space–time is not simply the presence of massive bodies in that region, but rather the associated distribution of **energy** and **momentum** throughout the region. The notion that particles of matter and radiation have energy will already be familiar, so the idea that a distribution of matter and radiation can be associated with a distribution of energy should not seem strange. The theory of special relativity, however, adds new depth to this idea since the relation $E = mc^2$ implies that even a stationary particle can be associated with a certain amount of energy. The idea of momentum may be less familiar, but in essence it is simply another physical quantity that, like energy, can be associated with any particle of matter or radiation once the mass and velocity of that particle are known. (Particles of radiation, such as photons, have zero mass but, according to special relativity, a photon of energy E carries momentum of magnitude $p = E/c$.) As far as cosmology is concerned, you have already seen that matter and radiation are spread throughout the Universe, so you should expect there to be some corresponding large-scale distribution of energy and momentum associated with all that matter and radiation. It is this large-scale distribution of energy and momentum, together with the equations of general relativity, which allow us to obtain a mathematical description of space–time curvature on the large scale. This is the basis of relativistic cosmology.

QUESTION 5.3

On the basis of what you learned in Section 5.2 about the large-scale distribution of matter and radiation, what word would you expect to characterize the large-scale distribution of the associated energy and momentum?

Although the discussion in this section has been very qualitative and general up to this point, some important ideas have been introduced, so it's worth summarizing them here.

- The important ingredients of the Universe include space and time as well as matter and radiation.
- Einstein's special theory of relativity taught us to regard the three-dimensional space and one-dimensional time with which we are familiar as a four-dimensional space–time, in which all matter and radiation is contained.

- Einstein's general theory of relativity taught us that space–time has geometrical properties (e.g. curvature) that are determined by the distribution of energy and momentum associated with matter and radiation.
- By combining our beliefs about the large-scale distribution of energy and momentum with the general theory of relativity it should be possible to obtain a mathematical model of space–time on a large scale.

The next section concerns the meaning of the phrase the 'geometric properties' of space–time, and the way in which those properties can be summarized mathematically. The section after that considers the large-scale distribution of energy and momentum. Having dealt with those two topics, the central equations of general relativity – Einstein's field equations – are introduced. We can then discuss the mathematical models of space–time that are consistent with our observations of matter and radiation, and the associated distribution of energy and momentum.

5.3.2 The space and time of the Universe

Imagine you were asked to describe space: not just the outer space beyond the atmosphere, but space in general, including the space you are occupying right now, and the space in which you might wave your arms without leaving your seat. You might say that space is big, or that it had three dimensions (i.e. three independent directions in which things can move), but what else could you do to describe space?

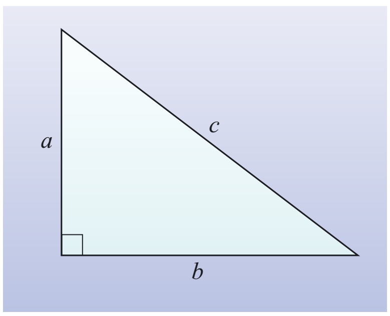


Figure 5.10 Pythagoras's theorem concerns the lengths, a, b and c, of the sides of a right-angled triangle (i.e. a closed, three-sided figure with one of its interior angles equal to 90°). In the notation of the figure, $c^2 = a^2 + b^2$.

The chances that you will be asked to describe space may be slim, but for a cosmologist the question is crucial and the conventional answer is well known. For cosmologists the description of space is essentially a matter of geometry. According to dictionaries, **geometry** is 'the study of the properties and relations of lines, surfaces and volumes in space'. It is by studying the properties and relations of objects *in* space that we learn about space itself. Now, geometry is a big subject, but 19th century mathematicians such as Carl Friedrich Gauss (1777–1855) and Bernhard Riemann (1826–1866) found powerful ways of summarizing the whole of geometry in just a line or two of mathematics. Gauss in particular, certainly one the greatest mathematicians who ever lived, initiated this development by recognizing the exceptional importance of **Pythagoras's theorem** about right-angled triangles.

According to Pythagoras's theorem, the square of the length of the longest side of a right-angled triangle is equal to the sum of the squares of the lengths of the other two sides. In symbols (see Figure 5.10), this can be expressed as

$$c^2 = a^2 + b^2 \tag{5.3}$$

Gauss realized that if this result held true for even the smallest conceivable right-angled triangles, then it could be used as the starting point for mathematical proofs of all the other known truths concerning the geometry of a two-dimensional plane. So, if we imagine an infinitesimally small version of Figure 5.10, and if we indicate its smallness by representing the lengths of its sides by ds, dx and dy rather than c, a and b, then we can say that the whole of two-dimensional plane geometry is implicitly contained in the single expression

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 \tag{5.4}$$

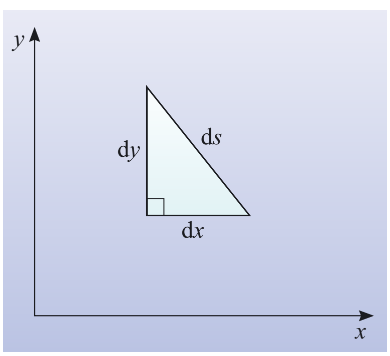


Figure 5.11 An infinitesimal version of Pythagoras's theorem.

where ds is the distance between two points whose x- and y-coordinates differ by the infinitesimal amounts dx and dy (see Figure 5.11). As far as a mathematician is concerned, Equation 5.4 is a complete answer to the question: 'Tell me all about the geometry of a two-dimensional plane.'

The form of Equation 5.4 immediately suggests an answer to the question 'Tell me all about the geometry of three-dimensional space.' If we describe any point in space by using the three-dimensional coordinate system with mutually perpendicular axes x, y and z (see Figure 5.12), then you can easily imagine that the distance $\mathrm{d}s$ between two points separated by infinitesimal coordinate differences $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$ is given by a three-dimensional generalization of Pythagoras's theorem:

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2 \tag{5.5}$$

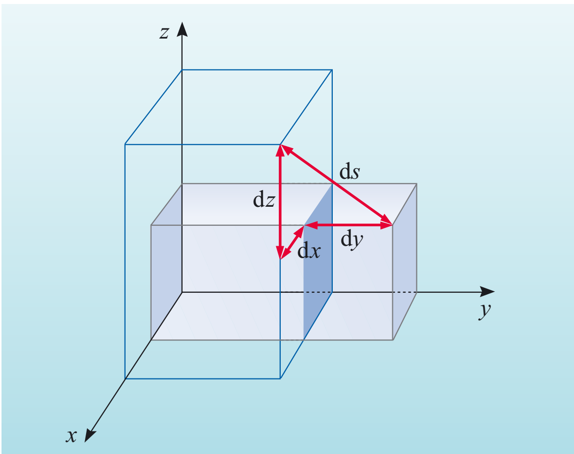


Figure 5.12 Two points in three-dimensional space with position coordinates that differ by the infinitesimal amounts $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$. The points are separated by a distance $\mathrm{d}s$.

This equation provides a basis for three-dimensional geometry, just as Equation 5.4 provides a basis for two-dimensional geometry. However, it's important to note that since Equation 5.4 only applies to a plane, it describes geometry on a *flat* surface. It does not, for example, describe the geometry of shapes drawn on the *curved* two-dimensional surfaces shown in Figure 5.13. Concise mathematical statements that summarize the geometry of these curved two-dimensional surfaces can be written down, but those equations are inevitably somewhat more complicated than Equation 5.4. Similarly, for all its power, Equation 5.5 only describes the geometry of what is confusingly called a 'flat' three-dimensional space. Mathematicians are familiar with similar but more complicated expressions that describe the geometry of 'curved' three-dimensional spaces, but we shall not write them down here. There is little point in trying to picture what a 'curved' three-dimensional space would be like, but it is worth emphasizing that in a flat space familiar geometric results hold true, while in a curved space those same results may cease to be true. For example in a flat space the interior angles of a triangle add up to 180°, but in a curved space this may not be true. Similarly, 'straight' lines that are initially parallel do not necessarily remain parallel in curved space.

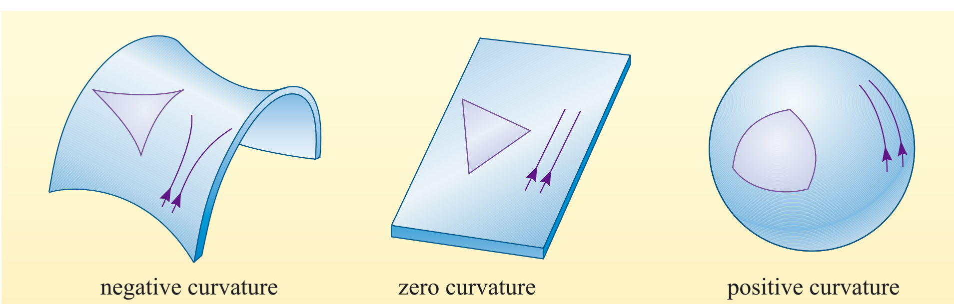


Figure 5.13 Some curved two-dimensional surfaces, viewed in three-dimensional space. The angles of triangles drawn on these surfaces do not add up to 180°, due to the curvature of the surfaces. Nor do lines that are initially parallel necessarily remain parallel.

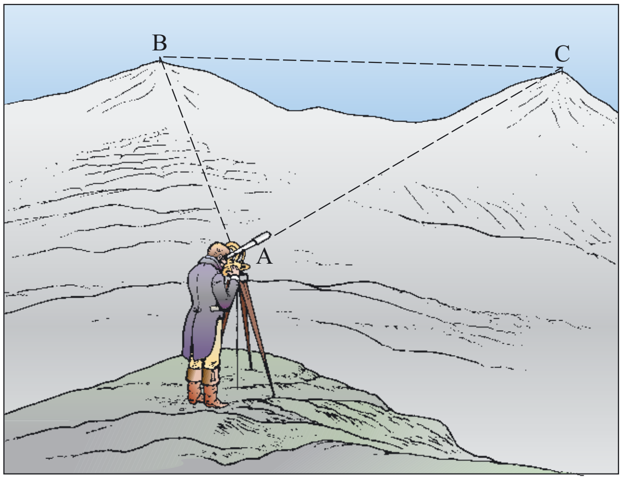


Figure 5.14 Gauss realized that mathematics provided many possible 'geometries' of space. The true geometry, flat or curved, could only be determined by experiment. He measured the interior angles of a triangle with vertices on three mountain tops, but found no deviation from 180° within the accuracy of his measurements. (Adapted from Kittel *et al*., 1965)

When Gauss realized that three-dimensional space might be curved he involved himself in a land survey that was being conducted, in order that he might investigate experimentally the properties of space (Figure 5.14). He hoped that accurate measurements of large triangles might reveal that their interior angles did not sum to 180°, implying that we are living in a curved space. (Note that Gauss was not concerned with triangles that followed the curvature of the Earth's surface; he was interested in the curvature of space itself – its *intrinsic* geometry – not the curvature of surfaces in space.) The results obtained in the survey did not provide any evidence that space is curved, but, as we now know in the light of Einstein's theory, that was simply because the curvature of space close to the Earth is too slight to be detected by the methods available to Gauss.

It has already been stressed that Einstein's theory of general relativity explains gravitational phenomena in terms of the curvature of space–time. So, you shouldn't be surprised that having discussed flat and curved three-dimensional spaces we now need to discuss the geometry of flat and curved four-dimensional space–times. Don't panic! As far as flat space–time is concerned all we need do is generalize Equation 5.5 so that instead of considering two narrowly separated *points* in space, we consider two neighbouring *events* in space–time. The two events can still differ in position by the infinitesimal amounts $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$, but, being events, we can also choose them so that the times at which they occur differ by the infinitesimal amount $\mathrm{d}t$. Investigations based on Einstein's theory of special relativity, which does not include the effects of gravity and therefore concerns flat space–time, show that the appropriate generalization of Equation 5.5 is

Events play the same role in space–time that *points* play in space. Whereas a point in space can be specified by three position coordinates (x, y, z), an event in space–time requires three position coordinates and a time coordinate (x, y, z, t). A point is an idealized location; an event is an idealized occurrence.

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2 - c^2(\mathrm{d}t)^2 \qquad (5.6)$$

where c is the speed of light in a vacuum, a fundamental physical constant that would have been 'discovered' in the theory of relativity even if it had not already been known from studies of light.

It would be inappropriate to spend time justifying the precise form of Equation 5.6, but, as before, you should realize that Equation 5.6 describes the geometry of a 'flat' (i.e. zero curvature) space–time. The corresponding equation for a curved space–time will inevitably be more complicated.

5.3.3 The distribution of energy and momentum in the Universe

As explained in Section 5.2.3, the observation that the Universe appears to be the same in all directions, combined with the Copernican principle – that we are not observing from a privileged position – implies that the Universe should be the same everywhere, on the large scale. As was also stated, the implied uniformity in the large-scale distribution of matter and radiation is now being increasingly well confirmed by observations, particularly the deep redshift surveys that are being carried out. Even so, because the observational evidence is necessarily limited, it is still appropriate to treat uniformity on the largest scales as an assumption supported by increasingly good evidence rather than a proven fact.

Cosmologists usually call this assumption of large-scale uniformity the **cosmological principle**, and sometimes state it in the following way.

> On sufficiently large size-scales (i.e. averaged over regions that are several hundred megaparsecs across) the Universe is **homogeneous** (i.e. the same everywhere) and **isotropic** (i.e. the same in all directions).

The technical terms 'homogeneous' and 'isotropic' make precise the rather loose notion of 'uniformity' that we have been using up to this point. Both terms are needed because it is possible for a distribution to be homogeneous without being isotropic. For instance, a universe in which there was a homogeneous magnetic field that everywhere pointed in the same direction would not be isotropic, though it would be homogeneous.

■ What is the precise feature of the cosmological principle that rules out the uniformly magnetized universe that has just been described? Explain your answer.

❑ Although the magnetized universe would be homogeneous it would not be isotropic. No point in the universe would be distinguished from any other point, as homogeneity demands, but at any point it would be possible to identify a 'preferred' direction by using a compass to determine the direction of the magnetic field. The fact that all directions are not equivalent in that universe shows that it is not consistent with the requirement for isotropy in our Universe.

In the simplest cosmological models that are consistent with the cosmological principle it is usually imagined that the Universe is completely filled with a uniform gas or fluid. (You can think of superclusters of galaxies as being the equivalent of 'atoms' in this cosmic fluid.) One advantage of taking such a simplified view of the contents of the Universe is that describing the state of the gas at any time t only involves specifying the *density* and *pressure* at the relevant time. These two properties of the gas determine all the other properties, such as the temperature. Density and pressure are usually denoted by the symbols ρ (the Greek letter 'rho') and p, but in an expanding Universe the density and pressure should be expected to change with time and we can indicate this dependence on time by writing the density and pressure at any time t as $\rho(t)$ and $p(t)$. Cosmological discussions are often further simplified by assuming that the pressure is negligible. This seems to be a reasonable assumption throughout much of cosmic history, since there is no evidence of superclusters colliding and rebounding in the way that atoms in a gas are supposed to do. We shall assume that pressure is negligible throughout most of this chapter but not in Chapter 6, which deals with the hot, dense, early Universe.

Thanks to the simplifying assumptions outlined above, it is quite easy for a cosmologist to write down a mathematical description of the large-scale distribution of energy and momentum in the Universe. Like the uniformly distributed cosmic gas, the energy and momentum have a homogeneous and isotropic distribution that can easily be described mathematically. Armed with this mathematical description of the energy–momentum distribution, cosmologists are able to use the equations of general relativity to determine the large-scale geometry of space–time and hence formulate a cosmological model. The first models of this kind are discussed in the next section.

5.3.4 The first relativistic models of the Universe

According to general relativity, the distribution of energy and momentum determines the geometric properties of space–time, and, in particular, its curvature. The precise nature of this relationship is specified by a set of equations called the *field equations* of general relativity. In this book you are not expected to solve or even manipulate these equations, but you do need to know something about them, particularly how they led to the introduction of a quantity known as the *cosmological constant*. For this purpose, Einstein's field equations are discussed in Box 5.1.

BOX 5.1 EINSTEIN'S FIELD EQUATIONS OF GENERAL RELATIVITY

When spelled out in detail the **Einstein field equations** are vast and complicated, but in the compact and powerful notation used by general relativists they can be written with deceptive simplicity. Using this modern notation, the field equations that Einstein introduced in 1916 can be written as

$$\boldsymbol{G} = \frac{-8\pi G}{c^4}\boldsymbol{T} \tag{5.7}$$

Different authors may use different conventions for these equations. The bold symbols $\boldsymbol{G}$ and $\boldsymbol{T}$ represent complicated mathematical entities called *tensors* that it would be inappropriate to explain here, except to say that $\boldsymbol{G}$ describes the curvature of space–time while $\boldsymbol{T}$ describes the distribution of energy and momentum, and that both these quantities may vary with time and position. The other symbols just represent numerical constants and have their usual meanings, $G = 6.673 \times 10^{-11}\ \mathrm{N\,m^2\,kg^{-2}}$ is Newton's gravitational constant and $c = 2.998 \times 10^8\ \mathrm{m\,s^{-1}}$ is the speed of light in a vacuum.

Given the distribution of momentum and energy at all points in space and time (i.e. given $\boldsymbol{T}$), the field equations determine the geometrical quantity $\boldsymbol{G}$, from which it may be possible to derive a detailed description of space–time geometry along the lines of the infinitesimal generalizations of Pythagoras's theorem that were discussed in Section 5.3.2. In his 1916 paper, Einstein used these equations to investigate planetary motion in the Solar System and to predict the non-Newtonian deflection of light by the Sun.

Einstein published his first paper on relativistic cosmology in the following year, 1917. In that paper he tried to use general relativity to describe the space–time geometry of the whole Universe, not just the Solar System. While working towards that paper he realized that one of the assumptions he had made in his 1916 paper was unnecessary and inappropriate in the broader context of cosmology. This led him to introduce another term into the field equations, a term he had deliberately chosen to ignore in 1916. With this extra term included the field equations used in the cosmology paper of 1917 can be written

$$\boldsymbol{G} + \Lambda\boldsymbol{g} = \frac{-8\pi G}{c^4}\boldsymbol{T} \tag{5.8}$$

As you can see, the extra term takes the form $\Lambda\boldsymbol{g}$, where Λ is the upper case Greek letter 'lambda'. The $\boldsymbol{g}$ here is another of these tensor quantities that was actually already implicitly involved in $\boldsymbol{G}$, while Λ represents a new physical constant called the **cosmological constant**. Provided Λ is sufficiently small (which it is) the presence of the $\Lambda\boldsymbol{g}$ term does not invalidate any of the results that Einstein obtained in the 1916 paper but, in the context of cosmological calculations, a positive value of Λ implies the action of a long-range repulsion that might, under appropriate circumstances, balance or even overwhelm the attractive influence of gravity.

In 1917 there was no evidence to indicate that the Universe was either expanding or contracting. So, in developing the first relativistic cosmological model, Einstein sought a value for the cosmological constant Λ that would ensure everything was constant and unchanging with time. He was also guided by the cosmological principle, so he required that the average density of matter in the Universe, ρ, should be homogeneous (i.e. independent of position) as well as constant (i.e. independent of time). He ignored the possibility of pressure, effectively assuming $p = 0$. Using these assumptions together with the modified field equations (Equation 5.8), Einstein constructed the first relativistic cosmological model, which is now

known as the **Einstein model**. The need to balance the gravitational attraction of matter and the repulsive effect of the cosmological constant led Einstein to the relation

$$\Lambda = \frac{4\pi G\rho}{c^2} \tag{5.9}$$

Because it is not expanding, the Einstein model is not thought to represent the Universe we actually inhabit. In fact, after the expansion of the Universe had been discovered, Einstein himself described his first use of the cosmological constant as his 'greatest blunder'. Nonetheless, it is worth exploring the geometrical properties of the Einstein model since it can provide insight into some of the extraordinary possibilities of relativistic cosmology.

The universe described by the Einstein model is **static**, neither expanding nor contracting. In this model universe, space is **finite**, having a total volume that is proportional to $\Lambda^{-3/2}$. Despite being finite, space in the Einstein model is also **unbounded**, that is to say, you can travel as far as you like in any direction without ever hitting a wall or encountering anything like an 'edge' of space. However, if you were to travel far enough in a straight line, you would eventually find yourself back at your starting point.

How is this possible? How can a straight line close back upon itself? Very simply, because what we are discussing here is a straight line in a *curved* space. In the Einstein model, which is homogeneous and isotropic, the curvature of space must be the same everywhere and in all directions. In addition, this uniform curvature has a positive value at every point, which means that a 'straight' line will be uniformly bent back upon itself all along its length. The actual value of the positive curvature depends on the value of the cosmological constant Λ, with the consequence that a line that is as straight as it can be in any region (the kind of line defined by a light ray, say) will close on itself after a distance that is proportional to $\Lambda^{-1/2}$. Figure 5.15 attempts to provide some idea of the geometry of space–time in the Einstein model.

If we did inhabit the kind of universe described by the Einstein model, we might expect astronomical observations to reveal distant images of the Earth, Sun or Milky Way, due to light that had travelled along the closed paths that the model implies.

- If we did live in the universe described by the Einstein model, how might we determine the value of the cosmological constant?

- By measuring the average density of matter on the large scale, ρ, and then using Equation 5.9 to determine Λ.

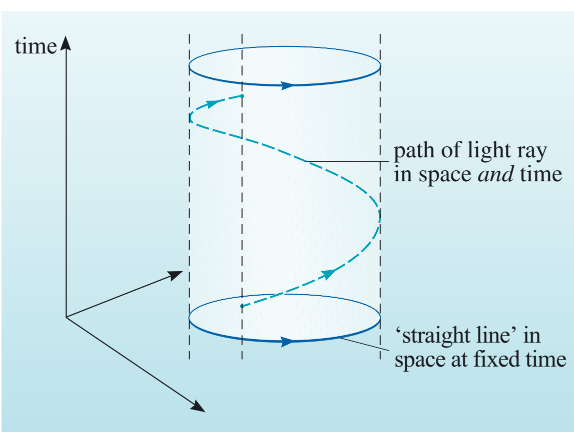


Figure 5.15 In this attempt to represent the four-dimensional space–time of the Einstein model, time is measured along the vertical axis. The whole of space at any time must therefore be represented in the horizontal plane, but because of the need to indicate that space is intrinsically curved, the circle that you see in the horizontal plane actually represents a 'straight' line. The helical path drawn above the circle might represent the path of a pulse of light that follows a 'straight' line through space while time passes. Because of the nature of this space–time diagram, the Einstein model is sometimes called 'Einstein's cylindrical world', but it's important to realize that the four-dimensional universe described by the Einstein model is no more akin to a real 'cylinder' than the 'flat' space of special relativity is akin to a pancake. (Adapted from Raine, 1981)

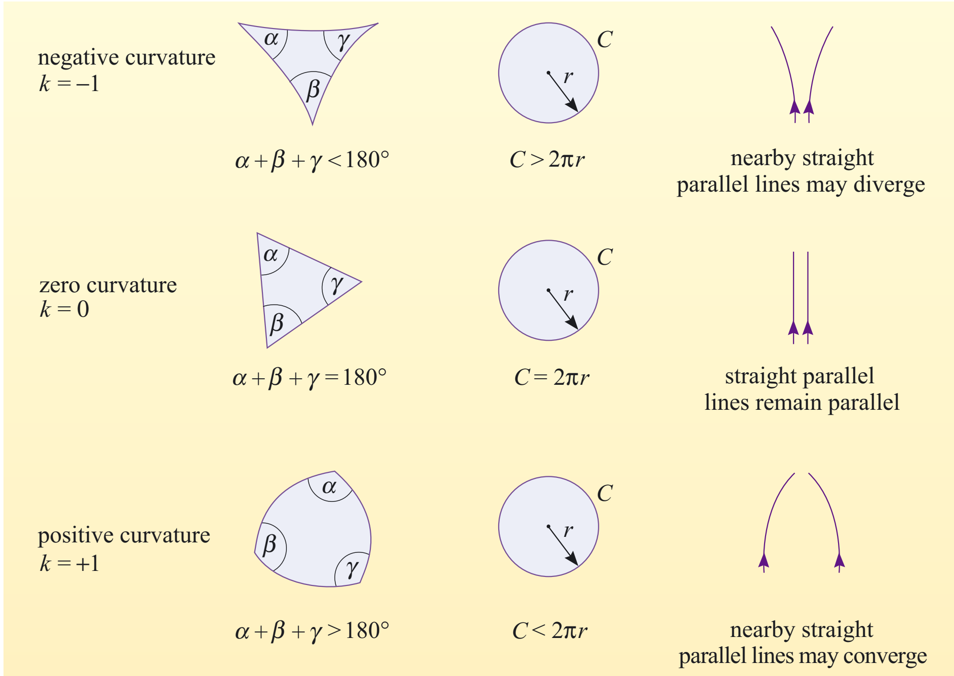


Figure 5.16 The effect of the curvature parameter, k, in determining the large-scale geometry of a cosmological model.

As you have seen, an important feature of the Einstein model is that the curvature of space is everywhere positive. This is conventionally indicated by introducing a **curvature parameter**, k, that may take the value +1, 0 or –1, and by saying that in the Einstein model $k = +1$. You will shortly be meeting other cosmological models with $k = 0$ and $k = -1$, as well as more models with $k = +1$. The curvature parameter is one of the most important characteristics of these models, since it strongly influences the large-scale geometric properties of the model. The value of k immediately determines whether space is finite ($k = +1$) or infinite ($k = 0$ or –1). As Figure 5.16 indicates, it also determines whether the interior angles of cosmically large triangles will add up to be less than, equal to or greater than 180°, how the circumference of a circle is related to its radius, and whether parallel pathways will remain parallel. The significance of this should soon become apparent, because the next model we are going to discuss has zero curvature everywhere and is characterized by $k = 0$. In this model, space is infinite, and 'straight' lines do not close back upon themselves.

Figure 5.17 Willem de Sitter was a Dutch astronomer who devised the second relativistic cosmology in association with his colleague Paul Eherenfest (1880–1933). He realized that, in his model, observers would find that the light from distant galaxies was red-shifted, but he described the outward radial velocity this indicated as 'spurious'. (Science Photo Library)

Within a year of Einstein's publication of the first relativistic cosmological model, a radically different model was proposed by the Dutch astronomer Willem de Sitter (1872–1934; Figure 5.17). As in the case of the Einstein model, the mathematical details of the **de Sitter model** can be found by solving the field equations (Equation 5.8). Like Einstein, de Sitter assumed that the Universe was homogeneous and isotropic, in accordance with the cosmological principle, but he did not require it to be static. Instead he took the view that the effect of matter was negligible, implying that both the average pressure and the average density could be taken to be zero. The geometric properties of space would therefore be determined by the cosmological constant alone. If positive, this would result in a never-ending expansion that would cause points in space to move apart perpetually. What little matter was actually present in the Universe would simply be carried along by the expansion of the space in which it was located (see Box 5.2). This is the key feature of de Sitter's model: it was the first model to describe an expanding Universe, although de Sitter himself seems not to have appreciated all the implications of this.

BOX 5.2 GALAXY RECESSION AND THE EXPANSION OF SPACE

At first sight, the observed recession of distant galaxies clearly suggests that those galaxies are moving through space. This, however, neglects the general relativistic view that space (or more properly space–time) has intrinsic properties that are influenced by the contents of the Universe. A useful way of thinking about space in general relativity is to imagine it as the three-dimensional analogue of a rubber sheet that may expand or contract as it is stretched or released.

If you picture galaxies as something like buttons placed on the rubber sheet, then it is clear that *one* way of causing them to separate is to move them across the sheet. This is the analogue of galaxies moving through space. But there is also a second way of increasing their separation. This other way of increasing their separation is to stretch the rubber itself. It is this latter view is that is most helpful when trying to interpret a phrase such as ‘matter would be carried along by the expansion of space’.

The galaxies we actually observe can be thought of as moving for two reasons. On the one hand they are being carried along *by* space as a result of the uniform cosmic expansion described by the Hubble flow. On the other hand those same galaxies are also moving *through* space due to local effects, such as the gravitational attraction of nearby concentrations of matter. The local effects can be dominant on the small scale, thus explaining why *some* nearby galaxies are moving towards us, but on larger scales such effects are so overwhelmed by cosmic expansion that *all* distant galaxies are found to be moving away from us.

One way of describing the expansion of space mathematically is to start with a coordinate grid that can expand or contract along with space. Such coordinates are said to be **co-moving** and are widely used in cosmology. This sort of coordinate system is indicated schematically in Figure 5.18, although for the sake of clarity only a two-dimensional grid is shown, rather than the three-dimensional grid that would really be required to label all the points in space. Due to the use of co-moving coordinates, typical points in an expanding space have unchanging coordinates, even though those points are moving apart. Since the coordinates themselves do not describe the expansion of space, it is necessary to introduce another parameter that does. This is called the **scale factor** and is represented by $R(t)$, where the parenthesized t indicates that the scale factor can change with time, increasing or decreasing as the Universe expands or contracts. If $R(t)$ increases with time, so that its value at time t_2 is greater than its value at some earlier time t_1, then the physical distance between two points with fixed coordinates will also increase, as ‘expansion’ implies that it should (this is the case shown in Figure 5.18). If $R(t)$ decreases with time then the distance between typical points is reduced, and space may be said to be contracting.

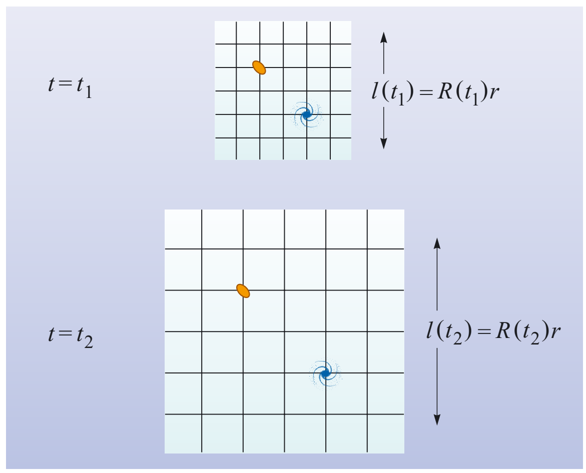


Figure 5.18 When co-moving coordinates are used to identify points, the expansion (or contraction) of space can be indicated by the behaviour of a scale factor, $R(t)$. If two points are separated by a co-moving coordinate distance r, the physical distance (in metres, say) between those two points, at time t, will be $l(t) = R(t)r$.

It’s important to realize that the use of co-moving coordinates removes any direct relationship between the coordinate differences dx, dy and dz of two narrowly separated points and the actual physical distance, ds, between those points. In order to find the real physical distance between the two points (measured in metres, say), at time t, the scale factor $R(t)$ must be taken into account. In the simplest case of an expanding flat space with zero curvature, this can be indicated by writing

$$(\mathrm{d}s)^2 = [R(t)]^2[(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2] \tag{5.10}$$

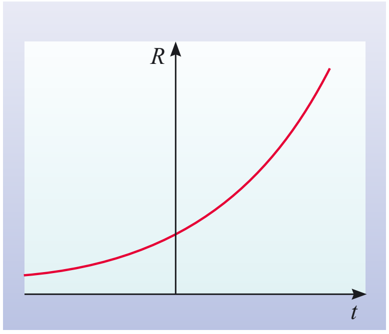


Figure 5.19 The behaviour of the scale factor in the de Sitter universe.

Equation 5.10 implies that if the 'coordinate distance' between the two points is dr, then the physical distance between them at time t is ds = $R(t)$ dr, where $\mathrm{d}r = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2}$.

In the case of the de Sitter model, the field equations show that the scale factor grows exponentially with time, as indicated in Figure 5.19. The steepness of the curve increases at a rate that is determined by the value of the cosmological constant, since it is Λ that drives the expansion. In fact, it can be shown that in the de Sitter model

$$R \propto \mathrm{e}^{Ht} \quad \text{where } H = \sqrt{\frac{\Lambda c^2}{3}} \tag{5.11}$$

Willem de Sitter realized that if matter moved in accordance with the expansion of space in his cosmological model, then distant astronomical bodies would be driven apart by the cosmic expansion. He also realized that this would give rise to red-shifts in the spectra of those bodies. He didn't press this point with sufficient vigour to deserve the credit for 'discovering' the expansion of the Universe, but Hubble was aware of this consequence of the de Sitter model and referred to it in his 1929 paper showing that redshift increased with distance. If the density of matter in the Universe had been negligible, and de Sitter's model had been correct, then the Hubble constant would have determined the value of H in Equation 5.11, and this would have allowed the value of Λ to be determined from the motion of distant galaxies.

QUESTION 5.4

The scale factor was not discussed in the context of the Einstein model but, if it had been, what could have been said about its behaviour?

5.3.5 The Friedmann–Robertson–Walker models of the Universe

Alexander Friedmann (Figure 5.20) was a Russian mathematician who worked at the University of St Petersburg. In 1922 and 1924 he published two important papers which had the effect of showing that the cosmological models of Einstein and de Sitter were special cases of a much wider class of models, all of which were consistent with the field equations of general relativity and with the cosmological principle. Later, Howard P. Robertson (Figure 5.21) and Arthur G. Walker (Figure 5.22) independently found improved ways of describing these models mathematically and ensuring their generality. It was the work of these three, Friedmann, Robertson and Walker, which resulted in the general mathematical framework that is still used today when discussing relativistic cosmological models of a homogeneous and isotropic Universe.

The geometric properties of space–time in any of the **Friedmann–Robertson–Walker models** (usually abbreviated to **FRW models**) can be deduced from the following expression for the space–time separation ds of two events whose coordinates differ by the infinitesimal amounts dx, dy, dz and dt, and which are located at a coordinate distance r from the origin.

There are several different but equivalent ways writing Equation 5.12. The form given here is not the most conventional.

$$(\mathrm{d}s)^2 = \frac{[R(t)]^2}{\left(1 + \dfrac{kr^2}{4}\right)^2}[(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2] - c^2(\mathrm{d}t)^2 \tag{5.12}$$

ALEXANDER FRIEDMANN (1888–1925), HOWARD PERCY ROBERTSON (1903–1961) AND ARTHUR GEOFFREY WALKER (1909–2001)

Figure 5.20 Alexander Friedmann. (Science Photo Library)

Figure 5.21 Howard Percy Robertson.

Figure 5.22 Arthur Geoffrey Walker.

Alexander Friedmann (Figure 5.20) was born and educated in St Petersburg and returned there in 1920 to work at the Academy of Sciences. Mainly known for his work on theoretical meteorology, Friedmann became interested in general relativity and used his mathematical talents to undertake a highly original exploration of the cosmological consequences of the theory. His researches led him to the equation that determines the evolution of the scale factor $R(t)$ in a homogeneous, isotropic universe that is uniformly filled with matter. This is now known as the Friedmann equation.

Howard Percy Robertson (Figure 5.21) was an American mathematical physicist who specialized in the application of general relativity to practical situations. In 1929, using general arguments that did not depend on specific assumptions about the properties of matter, Robertson deduced the general expression for the separation of events in the space–time of any universe that is spatially homogeneous and isotropic at all times (see Equation 5.12).

Arthur Geoffrey Walker (Figure 5.22) spent most of his academic career at the University of Liverpool, initially as a lecturer and later as Professor of Mathematics. His expression for the separation of events in a homogeneous and isotropic universe was published in 1936, and was based on a somewhat different approach from the earlier work of Robertson.

Equation 5.12 is sometimes described as the **Robertson–Walker metric**. We shall not be using this expression as the basis of any detailed arguments, but you should notice three things about it. First, it is a generalization (to the case of curved space–time) of Equation 5.6, which provided a complete description of the geometric properties of a flat space–time. Second, it contains the curvature parameter k, which helps to characterize the curvature of a space–time and can take the values -1, 0 or $+1$. Third, it contains the scale factor $R(t)$ that describes the expansion or contraction of space as a function of time.

Equation 5.12 applies to all the FRW models, but in order to work out the details of any particular model it is necessary to specify the value of k and to determine the precise form of $R(t)$. In the case of a universe uniformly filled with pressure-free (i.e. $p = 0$) matter of density ρ, the form of $R(t)$ can be determined by solving a complicated equation known as the **Friedmann equation** (Box 5.3). This important equation relates the value of R, and the rate of change of R, to the curvature parameter k, the cosmic density ρ and the cosmological constant Λ. We do not go into the details of its solution, but different values of k and Λ can lead to quite different forms for $R(t)$, and these are indicated schematically in Figure 5.23. The implications of these solutions are discussed below.

BOX 5.3 THE FRIEDMANN EQUATION: EPISODE 1

It should be clear from what has already been said that the Friedmann equation is of the utmost importance in the process of cosmological modelling. The Friedmann equation determines the form of $R(t)$, and thereby fixes the evolutionary history of a model universe. In this sense, the fate of the Universe is determined by the Friedmann equation, and many cosmologists would say that it is certainly the most important equation in cosmology.

All of this might make you wonder why we have not actually written down the Friedmann equation at this point. The reason is simple. The Friedmann equation involves mathematical notation and concepts with which you may not be familiar at this stage, but which arise naturally in the next section. We are therefore delaying the explicit introduction of the Friedmann equation until then.

If you really can't wait to see it, take a look at Episode 2 in Section 5.4.3 (in Box 5.4).

Figure 5.23 contains a great deal of information and deserves careful study. The first thing to notice is that each of the small graphs of R against t corresponds to different values (or ranges of values) of the curvature parameter k and the cosmological constant Λ. For each set of values, the small graph shows the history of spatial expansion or contraction in a homogeneous and isotropic universe filled with pressure-free matter. For example, the left-hand column (the column headed $\Lambda < 0$) contains all the cases where the cosmological constant is less than zero. The uppermost of the three graphs corresponds to $k = +1$, the middle graph corresponds to $k = 0$, and the lowest graph corresponds to $k = -1$. In all three of these cases the graphs are similar: R is 0 at $t = 0$, increases up to some maximum value, and then decreases to zero again after a finite time. In other words, all these models describe universes with a finite lifetime that expand, reach a state of maximum expansion, and then contract again. It is important to remember that the quantity R plotted in these graphs represents the 'scale' of the universe, not its radius. Only in the case where $k = +1$, implying that space has a finite total volume, does the concept of a 'radius' of the universe have any meaning; in the other cases, where $k = 0$ and $k = -1$, space is infinite and the concept of a cosmic 'radius' has no meaning. However, even these spatially infinite universes can expand and contract. By recalling that R is a *scale factor*, and that it characterizes the changing separation of typical (co-moving) points in a uniform universe, you will avoid making, or being misled by, meaningless statements about the radius or diameter of the universe.

The second column in Figure 5.23 is especially interesting; it contains the models that have a vanishing cosmological constant ($\Lambda = 0$). Until recently, these were believed to be the most realistic models of the Universe we actually inhabit. The first of the models in this class has $k = +1$, implying that space is finite, and the corresponding graph of R against t once again indicates a cycle of expansion and contraction with a finite lifetime. In this and other models that begin with $R = 0$ at time $t = 0$, the early part of the expansion is now known as the **big bang**; the collapse that takes place at the other end of the cycle is known as the **big crunch**. Among the $\Lambda = 0$ models, all start with a big bang but only the $k = +1$ model ends with a crunch: it is known as the **closed model**. In the other two models of the

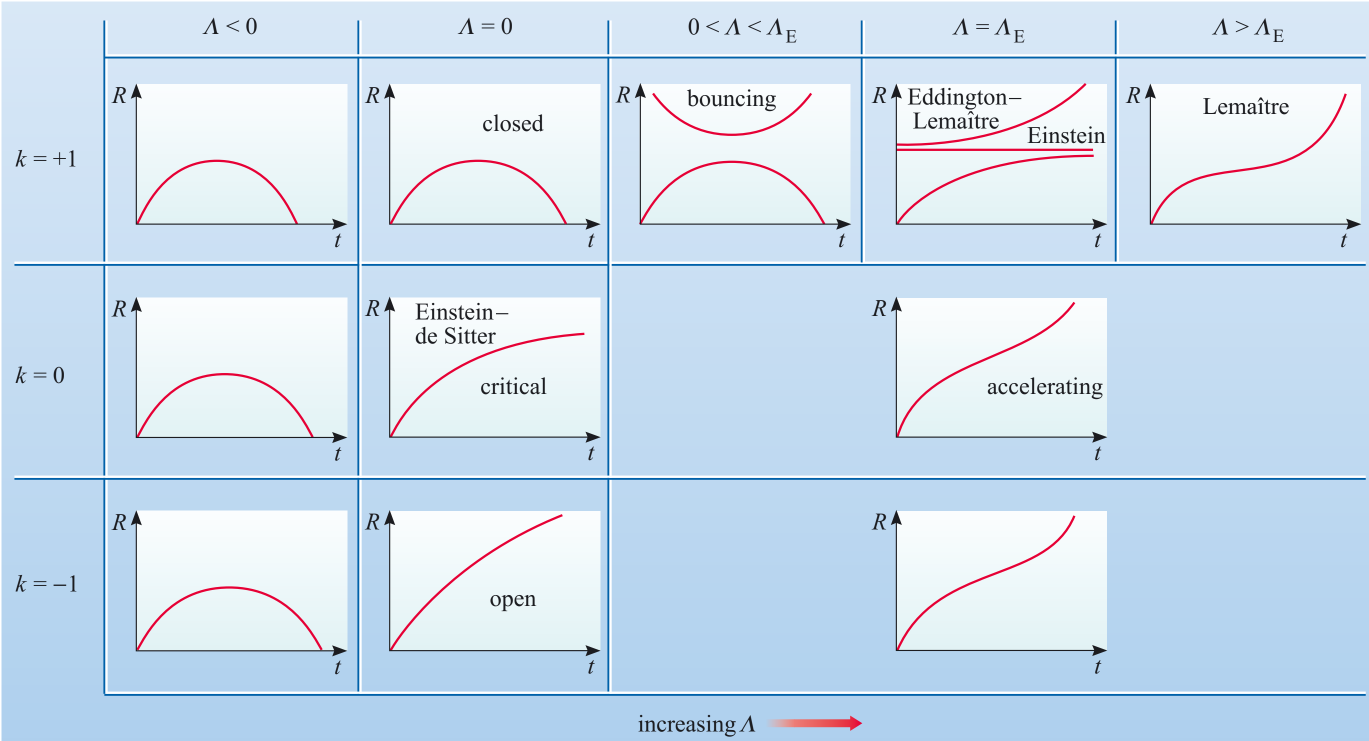


Figure 5.23 The Friedmann–Robertson–Walker models, classified according to the values of k and Λ. In each case the model is represented by a small graph of R against t, which encapsulates the history of spatial expansion and/or contraction implied by the model. Note that Λ_E represents the value of the cosmological constant in the Einstein model, $4\pi G\rho/c^2$. (Adapted from Landsberg and Evans, 1977)

$\Lambda = 0$ class, space is infinite and expands forever. The $k = -1$ model is called the **open model**. In this model, as t approaches infinity the relationship between R and t approaches the simple form $R \propto t$. The $k = 0$ model represents a special case between the open and closed models and is known as the **critical model**. In this case, the relationship between R and t takes the form $R \propto t^{2/3}$ for all values of t. Somewhat confusingly, the critical model is also known as the **Einstein–de Sitter model**, even though it has no direct relation to either the Einstein model or the de Sitter model.

In all of the remaining models of Figure 5.23 the cosmological constant is greater than zero ($\Lambda > 0$). It's best to regard all these models as occupying a single column, even though when $k = +1$ there are actually several quite distinct cases to discuss. But, before considering any of these models in detail, answer the following question.

- (a) In Figure 5.23, locate the graph of R against t that characterizes the Einstein model, and write down the corresponding values of k and Λ.

  (b) Can you see any graph in Figure 5.23 that corresponds to the de Sitter model?

- (a) The R against t graph for the Einstein model is the flat line shown in the middle of the top row of $\Lambda > 0$ models. According to Figure 5.23 it corresponds to $k = +1$ and $\Lambda = \Lambda_E$. (Note that Λ_E represents a particular value of the cosmological constant Λ. It follows from Equation 5.9 that, for a static universe of density ρ, that value is $\Lambda_E = 4\pi G\rho/c^2$.)

  (b) There is no graph in Figure 5.23 that is obviously identical to the R against t graph of the de Sitter model shown in Figure 5.19. However, as you will see below, the de Sitter model is present in Figure 5.23 as a 'limiting case' of the $\Lambda > 0$, $k = 0$ model.

Let's examine the $\Lambda > 0$ models in turn, starting with the case where $k = +1$ and Λ is greater than zero, but less than Λ_E (i.e. $0 < \Lambda < \Lambda_E$). In this case the graph indicates two possible kinds of behaviour. One is the now familiar situation in which R starts from zero at time $t = 0$, increases up to some maximum value and then decreases to zero again in a finite time. The alternative behaviour, indicated by the higher of the two curves, is one in which there is no obvious time to choose as $t = 0$, since there is no equivalent of the big bang. Rather, this is an infinitely old model in which an infinitely long period of contraction leads to a 'bounce' (when the scale factor reaches its minimum value) followed by an infinitely long period of expansion. If the Universe we actually inhabit is represented by a Friedmann–Robertson–Walker model with $k = +1$ and Λ in the range $0 < \Lambda < \Lambda_E$, then its actual behaviour – whether it follows the upper curve or the lower one – will be determined by its behaviour in the distant past. If the Universe really started with a big bang, the upper curve would be ruled out.

The next case to consider is that in which the cosmological constant has the particular value Λ_E that allows the model to be static. We have already noted that the static behaviour of the Einstein model is one of the allowed modes of behaviour in this case, and that is indicated by the presence of the flat line in the R against t graph. But other kinds of behaviour are also possible, as indicated by the other two curves in this part of the figure. One possibility, shown by the lower curve, is that R starts from zero and increases, gradually approaching the value specified by the Einstein model. The other possibility, represented by the upper curve, describes a universe that starts out in something very close to the static state, but expanding just a little. Even the tiniest initial expansion of this kind will eventually lead to a perpetual expansion, making it possible that an expanding universe might have had an indefinitely long history before the expansion really took hold, a possibility that many cosmologists have found attractive. This last kind of behaviour characterizes the **Eddington–Lemaître model**. The model was introduced in 1925 by Georges Lemaître (1894–1966; Figure 5.24), a Belgian cleric who made several contributions to relativistic cosmology, but it was elaborated and strongly advocated in a 1930 paper by the Cambridge astrophysicist Sir Arthur Eddington (1882–1944; Figure 5.25), who felt that it might well describe the real Universe.

Despite his pioneering work on the Eddington–Lemaître model, Georges Lemaître is particularly associated with the $k = +1$ model in which Λ is greater than Λ_E, which is known as the **Lemaître model**. Lemaître advocated this model in the 1930s, when a number of scientists became interested in the origin of the chemical elements, or, more specifically, the origin of the nuclei of the various elements. The Lemaître model describes a universe that is homogeneous and isotropic, in which space has a finite total volume at any time, and where R starts from zero at time $t = 0$ but increases without limit. In this model the expansion passes through a 'pseudo-static' phase in which the R against t graph becomes almost flat, so that, for a while at least, it resembles a static universe even though it is never truly static. Lemaître argued that the late stages of this model could represent the expanding Universe we currently observe, that the intermediate 'coasting' phase provided the necessary time for the formation of stars and galaxies, and that the early, highly compressed state would have been so hot and dense that it could account for the 'cooking up' of some of the nuclei that certainly exist in the Universe. Although the details of Lemaître's argument are no longer accepted, his notion that the first nuclei were formed in a hot, dense phase of the early Universe is widely accepted, and Lemaître is therefore generally credited with recognizing the importance of the 'big bang' even though he did not use that specific term.

GEORGES LEMAÎTRE (1894–1966) AND ARTHUR EDDINGTON (1882–1944)

Georges Lemaître (Figure 5.24) was a Belgian cosmologist who was also a Catholic priest. He initially trained as a civil engineer, but after serving in World War I he entered a seminary, became a priest, and subsequently studied solar physics in Cambridge. While there he met Arthur Eddington, and then visited America where he became familiar with the work of Hubble and Shapley. After returning to Belgium, Lemaître became Professor of Astrophysics at the University of Louvain in 1927. In 1931 he formulated his notion of an ultra-dense 'primeval atom', the explosion of which might start the observed expansion of the Universe.

Figure 5.24 Georges Lemaître. (Science Photo Library)

Figure 5.25 Sir Arthur Stanley Eddington. (Royal Astronomical Society Library)

Sir Arthur Stanley Eddington (Figure 5.25) was a Cambridge-based astrophysicist who made several important contributions to the study of stellar structure, particularly through his recognition of the significance of radiation pressure in maintaining (or destroying) equilibrium. Eddington was an early supporter of Einstein's general theory of relativity, and was the author of the first important book about the theory to appear in English. In 1919 he led the celebrated expedition that provided experimental support for the theory by observing the predicted deflection of starlight passing close to the edge of the Sun during a total eclipse.

In Figure 5.23, only one graph corresponds to $\Lambda > 0$ and $k = 0$. At the time of writing this is thought to be the model that most closely represents the real Universe. As you will see later, current astronomical evidence increasingly favours models in which $k = 0$. Observational evidence also favours $\Lambda > 0$. Hence this model represents the 'best buy' at the present time. As you can see from the R against t graph, this model represents a uniform universe that starts with a big bang and goes on expanding forever. The expansion undergoes a 'slowdown' at some stage, but does not exhibit the sort of 'pseudo-static' behaviour seen in the Lemaître model. After the slowdown, the rate of expansion increases continuously, and for this reason the model is sometimes referred to as the **accelerating model**.

The final graph in Figure 5.23 corresponds to $\Lambda > 0$ and $k = -1$. In a uniform universe of this type, space is infinite and has the kind of negatively curved geometry that causes cosmically large triangles to have interior angles that sum to less than 180°. This is another model that starts with a big bang. As in the accelerating model, the scale factor grows from zero, slows its growth temporarily and then accelerates again.

We have now discussed each of the FRW models, but we have still not found the de Sitter model among them. This is rather surprising, since Figure 5.23 should contain *all* the homogeneous and isotropic models that are filled by a pressure-free fluid. Where is the de Sitter model in this family? Well, the de Sitter model has $\Lambda > 0$ and $k = 0$, so we might expect to find it in the box that contains the accelerating model, and in fact that is where it is, but it is only present as a 'limiting case' of the behaviour that is illustrated in that part of the figure. The de Sitter model has a negligible amount of matter, so it corresponds to $\rho = 0$ as well as $p = 0$, whereas the graph that represents the $\Lambda > 0$ and $k = 0$ model in Figure 5.23 shows the

general case in which ρ may have a non-zero value. In this general case the density of matter decreases with time. As t increases the matter is eventually so thinly spread that such a universe increasingly resembles an essentially empty de Sitter model in which the cosmological constant is solely responsible for the expansion. As a result, the graph of R against t increasingly approaches the de Sitter form ($R \propto e^{Ht}$) as t approaches infinity. So, the de Sitter model is implicitly present in Figure 5.23, as a limiting case of what is shown. Though we shall not bother to discuss them, there are similar limiting cases elsewhere in Figure 5.23.

QUESTION 5.5

In the context of the Friedmann–Robertson–Walker models, which values or ranges of the parameters k and Λ correspond to universes with the following characteristics?

(a) The universe is neither homogeneous nor isotropic.

(b) There is no possibility of a big bang.

(c) A big bang is possible, but there is at least one other possibility (assume $\rho > 0$).

(d) The particular point in space where the big bang happened can still be determined long after the event.

(e) At any time, the large-scale geometrical properties of space are identical to those of a three-dimensional space with a flat geometry.

(f) Space has a finite volume, and 'straight' lines that are initially parallel may eventually meet.

(g) There is a big bang, but the volume of space is infinite from the earliest times.

5.4 The key parameters of the Universe

At the beginning of the last section it was stated that a cosmological model typically consists of:

- *equations* that imply general relations between observable quantities, together with
- *parameters* that must be determined by observation before the model can be used to provide detailed quantitative predictions.

You should fully appreciate the significance of this assertion now that you have examined the class of Friedmann–Robertson–Walker (FRW) models. You have just seen that in those models there is a general expression (Equation 5.12) that describes the geometry of space–time in terms of the separation of events. The form of this equation is enough to show a cosmologist that the universe being described is homogeneous and isotropic. However, a detailed appreciation of the properties of space–time in such a universe involves determining the parameters that arise in the model, specifically the curvature parameter k and a scale factor $R(t)$. Only when these are known does it become possible to evaluate quantities such as the curvature of space, which is determined at time t by the quantity $k/[R(t)]^2$. The importance of observable parameters is further emphasized by recalling that the behaviour of the scale factor is determined by the Friedmann equation, which involves the value of the curvature parameter k, the cosmological constant Λ and the average density of matter ρ, all of which are, in principle, observable parameters at any given time.

This section is concerned with the parameters that arise in the FRW models (basically k and $R(t)$), and their relationship to the observational parameters (such as the Hubble constant) that characterize our Universe. By determining and exploiting these relationships, it should be possible to use astronomical observations to determine which of the many cosmological models most closely resembles the Universe in which we live. This is one of the central challenges of the branch of cosmology known as *observational cosmology*.

Although this section concerns measurable parameters, its emphasis is on relationships rather than values. The values of the observable parameters, and the best ways of determining those values, are discussed much more fully in Chapter 7, which is entirely devoted to the subject of observational cosmology.

5.4.1 Hubble's law, the Hubble constant and the Hubble parameter

One observational result that finds a very natural explanation in the context of FRW cosmology is Hubble's law. You will recall that this law describes the general tendency for the redshift z of a galaxy to increase in proportion to its distance d from the observer, as described by the equation

$$z = \frac{H_0}{c} d \tag{5.1}$$

where the constant of proportionality, H_0/c, is made up of *Hubble's constant*, H_0, and the speed of light in a vacuum, c. You will also recall that for any particular galaxy the redshift z in Equation 5.1 can be related to the observed and emitted wavelengths, λ_{obs} and λ_{em}, of some identified spectral line by the equation

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{5.2}$$

By measuring the redshifts of distant galaxies, and independently measuring the distances of those galaxies, it is possible to use Equation 5.1 to determine the value of the Hubble constant H_0, since the speed of light is well known. Hubble himself did this, although his result was wildly inaccurate due to poor and incorrectly interpreted data. More modern observations have allowed the value of H_0 to be determined with an uncertainty of about 10% and there are good prospects of reducing this uncertainty still further. A recent estimate of the Hubble constant suggests a value of

$$H_0 = (2.3 \pm 0.3) \times 10^{-18}\,\mathrm{s}^{-1}$$

although, for historical reasons, this is more usually expressed in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (i.e. kilometres per second per megaparsec) giving

$$H_0 = (72 \pm 8)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

H_0 is one of the most important observational parameters in cosmology, but how does it relate to the FRW models, where the obvious parameters are k and $R(t)$, and there is no H_0 to be seen? This is what we must now investigate.

Figure 5.26 indicates the basis of the relationship. The figure shows two snapshots of an expanding FRW universe, with a growing scale factor $R(t)$. Two galaxies, A and B, happen to be located at the grid points of a set of co-moving coordinates that expands with the universe. The first snapshot represents a time t_{em} at which some light is emitted from galaxy A, and the second snapshot represents a later

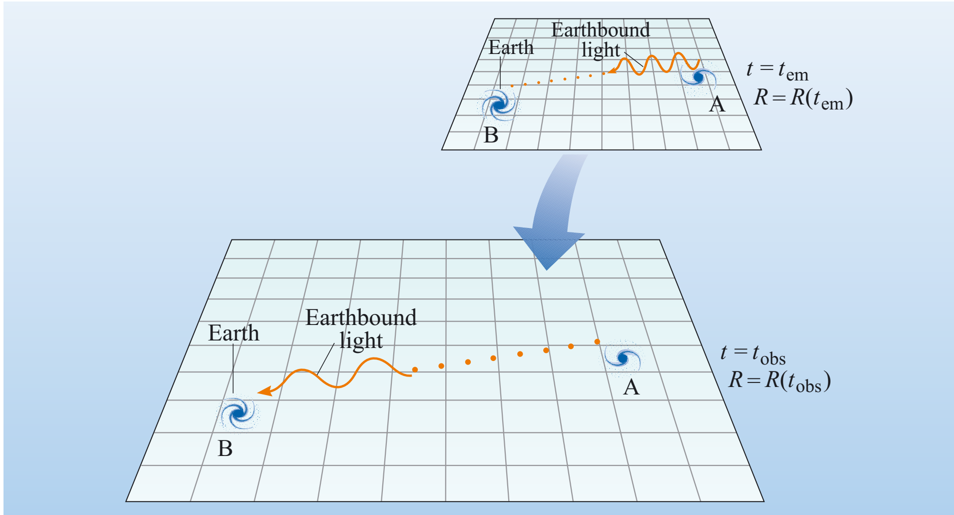


Figure 5.26 Cosmic expansion, measured by the increasing scale factor $R(t)$, as a cause of the cosmological red-shift of distant galaxies. The 'stretching' of space also stretches light waves travelling from one galaxy to another, causing the observed light to be red-shifted relative to the emitted light. (Adapted from Finkbeiner, 1998)

time t_{obs} at which that same light is observed in galaxy B. While the light is travelling from A to B, the (co-moving) coordinates of the galaxies do not change, but the physical distance between the galaxies does increase because it is proportional to $R(t)$, and $R(t_{obs})$ is greater than $R(t_{em})$. Whatever the distance from A to B at time t_{em}, it will have increased by a factor of $R(t_{obs})/R(t_{em})$ by the later time t_{obs}. Now, this expansion factor $R(t_{obs})/R(t_{em})$ represents the growth of space itself, so it will also influence the wavelength of the light that is moving freely between the two galaxies. As a result, light that is emitted from A at time t_{em} with wavelength λ_{em} will be observed at B at time t_{obs} with the longer wavelength $\lambda_{obs} = \lambda_{em} \times R(t_{obs})/R(t_{em})$. This increase in wavelength will, of course, be seen as a redshift by the observer in galaxy B.

- According to an observer in galaxy B at time t_{obs}, what is the redshift of galaxy A? Express your answer in terms of the expansion factor $R(t_{obs})/R(t_{em})$.

- Rearranging Equation 5.2

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{\lambda_{em}[(\lambda_{obs}/\lambda_{em}) - 1]}{\lambda_{em}}$$

Cancelling the overall factors of λ_{em}

$$z = \frac{\lambda_{obs}}{\lambda_{em}} - 1$$

Replacing $\lambda_{obs}/\lambda_{em}$ by the equivalent expansion factor $R(t_{obs})/R(t_{em})$, we find that

$$z = \frac{R(t_{obs})}{R(t_{em})} - 1 \qquad (5.13)$$

Note that according to the FRW model, the red-shift of a distant galaxy is primarily caused by the expansion of space, it is *not* a Doppler shift due to movement through space. These expansion-based redshifts are usually referred to as **cosmological redshifts** in order to distinguish them from Doppler shifts. Of course, real galaxies do not necessarily behave like the idealized galaxies of Figure 5.26. Real galaxies may have some 'peculiar' motion of their own relative to the

grid of co-moving coordinates, and this peculiar motion can give rise to Doppler shifts that cause the observed redshifts of galaxies to differ somewhat from the cosmological redshifts implied by a smoothly expanding FRW model.

We have now seen how redshifts can arise from expansion in a Friedmann–Robertson–Walker model, but the real point of Hubble's law is that the redshift z of distant galaxies *increases in proportion to their distance*. How do FRW models account for this? Very simply as it turns out. The greater the distance of a galaxy, the longer the light takes to reach us from that galaxy. The greater the time the light spends 'in flight' between the moments of emission and observation, the greater is the expansion factor $R(t_{\mathrm{obs}})/R(t_{\mathrm{em}})$, and the greater the redshift of the light, $z = [R(t_{\mathrm{obs}})/R(t_{\mathrm{em}})] - 1$.

The time-of-flight argument provides a qualitative explanation of Hubble's law in an expanding FRW model, but the explicit nature of the FRW models allows us to be even more precise about the exact nature of the relationship. In particular, it is possible to derive an equation that relates Hubble's constant to the scale factor. To see this, consider two galaxies separated by a relatively *small* distance d at time t when the scale parameter is $R(t)$. Because these galaxies are close together, the flight-time for light passing from one to the other, d/c, is also small and is represented by the quantity Δt. It follows from Equation 5.13 that the observed redshift of one of these galaxies, when observed from the other, is

$$z = \frac{R(t + \Delta t)}{R(t)} - 1 \qquad (5.14)$$

Due to the expansion of the Universe, $R(t + \Delta t)$ is greater than $R(t)$, and we can indicate this by writing $R(t + \Delta t) = R(t) + \Delta R(t)$, where the new quantity $\Delta R(t)$ represents the small increase in scale factor that occurs during the short time Δt. Note that $\Delta R(t)$ represents a single quantity, it is *not* the result of multiplying together quantities such as Δ and $R(t)$.

Replacing $R(t + \Delta t)$ in Equation 5.14 by the alternative expression $R(t) + \Delta R(t)$, shows that

$$z = \frac{R(t) + \Delta R(t)}{R(t)} - 1 \qquad (5.15)$$

and this can be rewritten as

$$z = 1 + \frac{\Delta R(t)}{R(t)} - 1 \qquad (5.16)$$

that is, $$z = \frac{\Delta R(t)}{R(t)} \qquad (5.17)$$

Now for the crucial step: $\Delta R(t)$ – the change in scale factor that occurs during the short time interval Δt – will be equal to the result of multiplying the interval Δt by the rate of change of R at the time t. (This is like saying that during a time Δt a car travelling with velocity v will change its position by an amount $\Delta t \times v$, since v is the rate of change of position.) It is conventional to represent the rate of change of the scale factor at time t by the symbol $\dot{R}(t)$ (read as 'R dot at time t'), so $\Delta R(t) = \Delta t \times \dot{R}(t)$, and to rewrite Equation 5.17 as:

$$z = \frac{\Delta t \times \dot{R}(t)}{R(t)} \qquad (5.18)$$

If you are familiar with differential calculus, you may find it helpful to know that $\dot{R}(t)$ is a common shorthand for the derivative $\mathrm{d}R(t)/\mathrm{d}t$.

Since $c/c = 1$, we can rewrite this as

$$z = \frac{c\Delta t}{c} \times \frac{\dot{R}(t)}{R(t)} \tag{5.19}$$

But, $c\Delta t = d$, the distance between the two galaxies. Using this, we can write

$$z = \frac{1}{c} \times \frac{\dot{R}(t)}{R(t)} \times d \tag{5.20}$$

Now, you should be able to see that this is similar to Hubble's law, according to which, at the present time t_0

$$z = \frac{H_0}{c} d \tag{5.1}$$

This similarity suggests that we should identify the time dependent quantity $\dot{R}(t)/R(t)$ in Equation 5.20 as a time dependent **Hubble parameter** that we can denote $H(t)$. Thus,

$$H(t) = \frac{\dot{R}(t)}{R(t)} \tag{5.21}$$

The value of this Hubble parameter varies with time, but the precise way that it varies depends on the precise way in which $R(t)$ varies with time, and will therefore differ from one FRW model to another. However, in any model that provides a good description of the real Universe, we expect to find that if we evaluate the 'theoretical' Hubble parameter at the present time t_0, then the value obtained should equal that of the observed Hubble constant, that is

$$H(t_0) = \frac{\dot{R}(t_0)}{R(t_0)} = H_0 \tag{5.22}$$

As you can see, we have now managed to relate an observational parameter, H_0, to the scale factor $R(t)$ – a parameter in the FRW model.

The function $\dot{R}(t)$ represents the rate of change of R at time t, so $\dot{R}(t)/R(t)$ represents the 'fractional' rate of change of the scale factor. Using this terminology we can say that in Friedmann–Robertson–Walker cosmology, the Hubble constant represents the fractional rate of change of the scale factor evaluated at the present time, t_0. More succinctly, we can say that the observed Hubble constant represents the current value of the model's Hubble parameter.

QUESTION 5.6

Figure 5.27 is a plot of redshift against distance for a number of galaxies. The plot includes a 'best fit' line drawn through the data. Assuming that our Universe can be well represented by an expanding Friedmann–Robertson–Walker model, state the significance of the gradient of the line, evaluate that gradient from the graph, and hence determine the value of the Hubble constant.

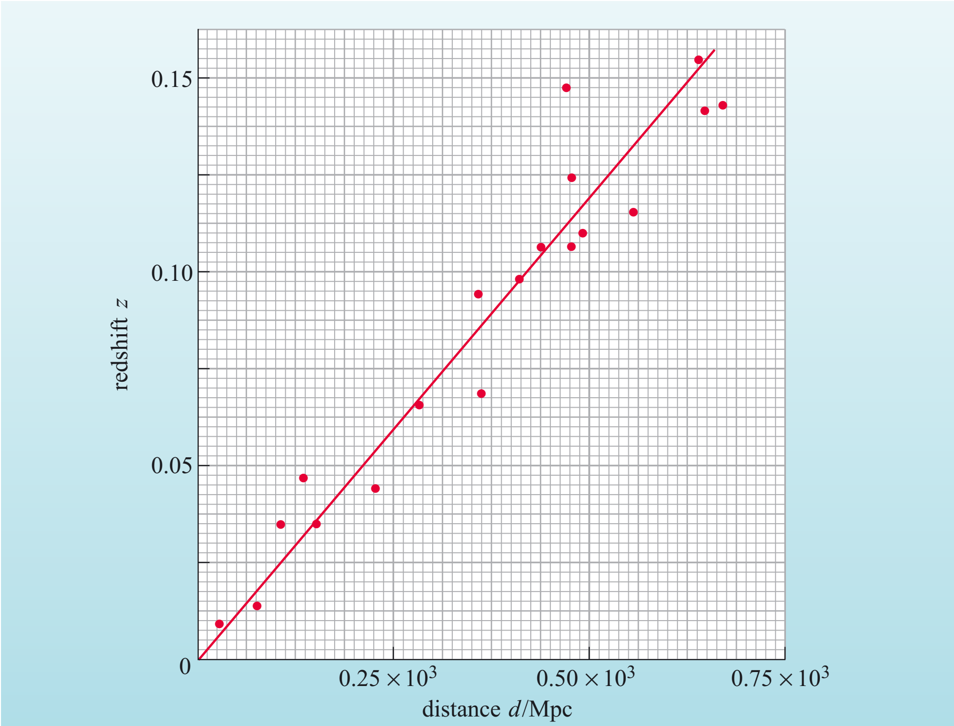


Figure 5.27 A plot of redshift against distance for a number of galaxies, together with a best-fit line through the data.

QUESTION 5.7

What is the rate of change of R in the Einstein model? What does your answer imply about the Hubble parameter in the Einstein model?

5.4.2 Systematic deviations from Hubble's law, and the deceleration parameter

At any time t, the Hubble parameter measures the rate expansion of the Universe and depends on the rate of change of R, which is indicated by the slope of the R against t graph. However, in most FRW models the expansion either speeds up or slows down as time progresses – that is to say the expansion is accelerating or decelerating – and this is indicated by the curvature of the R against t graph (see Figures 5.23 and 5.28).

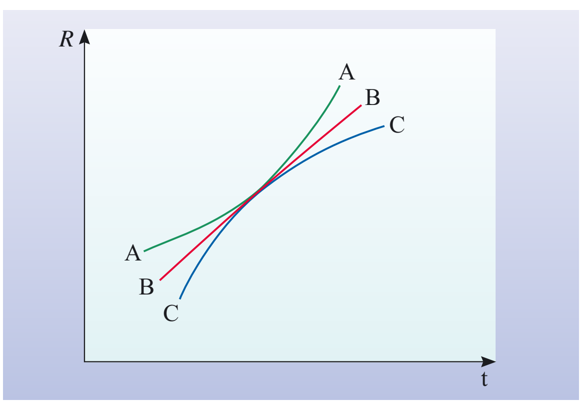


Figure 5.28 A plot of R against t, for large values of t, in three different FRW models denoted A, B and C. The upward curvature of A indicates acceleration, the downward curvature of C indicates deceleration and the relative lack of curvature in B indicates an almost steady expansion.

- Referring back to Figure 5.23 (and the accompanying text), identify three named FRW models that might correspond to the curves A, B and C in Figure 5.28.

- A could represent the accelerating model ($k = 0$, $\Lambda > 0$)
  B could represent the open model ($k = -1$, $\Lambda = 0$)
  C could represent the critical model ($k = 0$, $\Lambda = 0$).

Just as the rate of change of R is indicated by $\dot{R}(t)$, so the rate of change of $\dot{R}(t)$ is indicated by $\ddot{R}(t)$ (read as 'R double dot at time t'). If the expansion of the Universe is speeding up at time t then $\ddot{R}(t)$ will be positive, if the expansion is slowing down, $\ddot{R}(t)$ will be negative, and if there is no acceleration or deceleration $\ddot{R}(t) = 0$.

In the context of the FRW models, it turns out that a useful way of characterizing an increasing or decreasing rate of expansion is in terms of a quantity called the **deceleration parameter**. This varies with time, and is conventionally denoted by the symbol $q(t)$. It is defined as follows

$$q(t) = \frac{-R(t)}{[\dot{R}(t)]^2}\ddot{R}(t) \tag{5.23}$$

Note the negative sign in this equation; this implies that if $\ddot{R}(t)$ is positive (i.e. if the expansion is accelerating) then the deceleration parameter will be negative. In the cases shown in Figure 5.28, at large values of t the deceleration parameter would be negative for curve A, zero for curve B, and positive for curve C.

- What would you expect the corresponding results to be for small values of t in the three FRW models named in the last question?

- In all three models the R against t graph curves downwards at early times, similar to the behaviour of the Einstein–de Sitter model at very early times. This indicates that in its early phases the expansion is decelerating, and implies that $q(t)$ will be positive in all three cases.

Now, the detailed argument presented in the last section showed that the expanding FRW models predict a direct proportionality between z and d of just the kind described by Hubble's law. However, that argument was based on the behaviour of galaxies that were sufficiently close together for the time of flight of light passing from one to the other to be considered 'short'. When more distant galaxies are taken into account, the FRW models predict that the direct proportionality between z and d will break down, and systematic deviations from Hubble's law will be observed. Moreover, the models show that the systematic deviations from Hubble's law depend on the value of the deceleration parameter at the time of observation.

In fact, the FRW models predict that present-day observations of galaxy redshifts and distances should show, to a first approximation, that

$$d = \frac{cz}{H_0} \tag{5.24}$$

which is just a rearrangement of Hubble's law, and agrees with observations for redshifts of less than 0.2 or so. But the FRW models also predict that, to a better approximation

$$d = \frac{cz}{H_0}\left[1 + \frac{1}{2}(1 - q_0)z\right] \tag{5.25}$$

where q_0 is the current value of the deceleration parameter. Figure 5.29 illustrates this relationship by showing the kind of systematic deviations from Hubble's law that might be expected for various values of q_0 out to a redshift of about 1. It's worth noting that the distance d in these relationships represents the distance indicated by an object's luminosity at the time of observation, t_0.

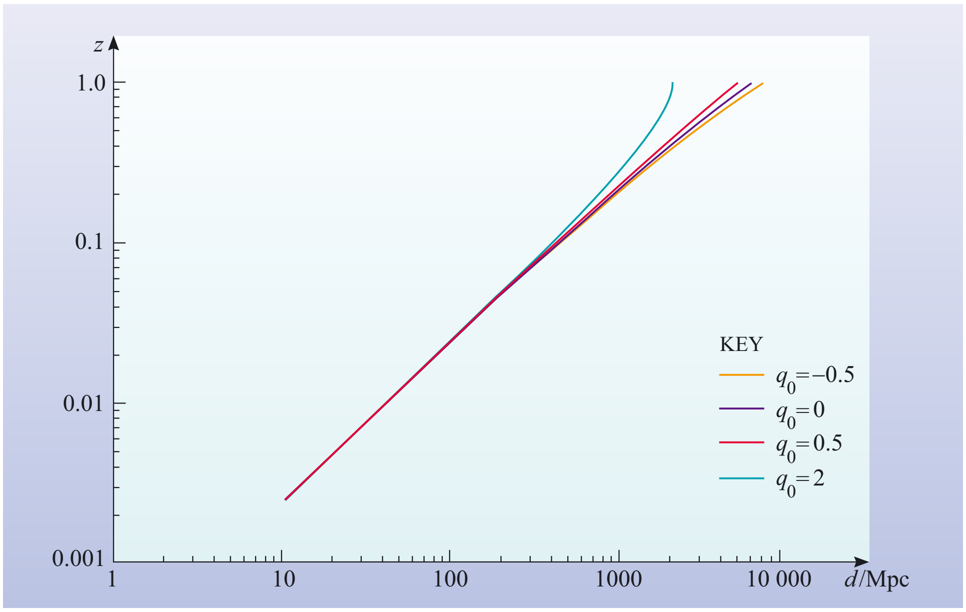


Figure 5.29 The current form of the graph of redshift against distance for galaxies, as predicted by expanding FRW models. The precise shape of the curve depends on the current value of the deceleration parameter, q_0, but all such models predict a straight part that is determined by the current value of the Hubble parameter, H_0. (In this case $H_0 = 72\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ has been assumed.)

In principle then, given sufficiently good observational data, the straight part of the redshift against distance graph can be used to determine the current value of the Hubble parameter, H_0, and the observed deviations from straightness can be used to determine the current value of the deceleration parameter, q_0. The determination of these two values, H_0 and q_0, was, for many years, the primary objective of observational cosmology. In fact, the American astronomer Alan Sandage (1926–), a former assistant of Hubble, once famously described observational cosmology as 'the search for two numbers', a characterization that was largely true until the 1970s.

Although there is now far more to observational cosmology than the determination of H_0 and q_0, the determination of those two numbers is still of very great importance. As mentioned earlier, recent observations have reduced the uncertainty in the value of H_0 to about 10% and even greater accuracy is to be expected soon. However, the determination of q_0 presents a much greater challenge. In order for the deviations from Hubble's law to be seen it is necessary to observe galaxies at large redshifts and to independently measure their distances. Finding high redshift galaxies is now relatively easy, but accurately determining their distances is very hard. A number of recent observations (see Chapter 7 for details) have indicated that q_0 is negative, implying that the expansion of the Universe is accelerating. If these results hold up, and if we do indeed live in a Universe broadly described by an accelerating FRW model, then it must also be the case that the cosmological constant is greater than zero.

QUESTION 5.8

Justify the assertion made in the last sentence above.

5.4.3 The critical density and the density parameters

An important parameter in any FRW model is the average density of cosmic matter. In an expanding or contracting Universe this quantity will change with time, so in the context of FRW models it is represented by the symbol $\rho(t)$. Observationally we might hope to determine the current value of the density, $\rho(t_0)$, by adding together the masses of all the galaxies and clusters in some sufficiently large region of the Universe, and dividing that sum by the volume of the region. Of course, this has been attempted many times, but the answers are fraught with uncertainties, partly due to the problems of measuring large distances and observing faint galaxies, but also due to the very great problem of determining the total mass of dark matter in any region. For this reason, more indirect approaches to the determination of ρ are usually necessary.

When discussing the cosmic density a useful reference value is the density of matter in the FRW model with $k = 0$ and $\Lambda = 0$. You will recall from our discussion of Figure 5.23 that this particular model is often referred to as the *critical model*, since it sits exactly on the borderline between the open and closed $\Lambda = 0$ models. It turns out that, in order to maintain this precarious position, the density of matter in the critical model must be precisely related to the value of the Hubble parameter at all times. In fact, if we denote the density of matter in the critical model at time t by $\rho_{\text{crit}}(t)$, the Friedmann equation (see Box 5.4) implies that

$$\rho_{\text{crit}}(t) = \frac{3H^2(t)}{8\pi G} \tag{5.26}$$

where G is Newton's gravitational constant. The quantity $\rho_{\text{crit}}(t)$ is known as the **critical density** at time t. Note that both $\rho_{\text{crit}}(t)$ and $H(t)$ vary with time, but, in the critical model, Equation 5.26 always relates their variations. So, it is always possible to work out the current value of the critical density from the current value of the Hubble parameter.

■ Until recently it was widely believed that our Universe was well represented by the critical model. If this belief had been correct what would have been a reasonable estimate of the current value for the density of the Universe?

❑ Under these conditions the current value of the density, $\rho(t_0)$, would be the current value of the critical density, i.e. $3H_0^2/8\pi G$. Using the value for H_0 given in Section 5.4.1, i.e. $72\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1} = 2.3 \times 10^{-18}\,\text{s}^{-1}$, we see that

$$\rho_{\text{crit}}(t_0) = \frac{3 \times (2.3 \times 10^{-18})^2}{8 \times \pi \times 6.67 \times 10^{-11}}\,\text{kg}\,\text{m}^{-3} \approx 1 \times 10^{-26}\,\text{kg}\,\text{m}^{-3}$$

Using the critical density as a reference value, we can express the actual density of cosmic matter at any time as a fraction of the critical density at that time. This fraction is called the **density parameter for matter**, and may be represented by the symbol $\Omega_{\text{m}}(t)$, so

Ω is the upper case Greek letter 'omega'.

BOX 5.4 THE FRIEDMANN EQUATION: EPISODE 2

This is a good place to finally write down the explicit form of the crucially important Friedmann equation that was first discussed in Section 5.3.5. The equation makes use of the symbol $\dot{R}(t)$ to represent the rate of change of the scale factor R, and may be written as

$$\dot{R}^2 = \frac{8\pi G R^2}{3}\left(\rho + \frac{\Lambda c^2}{8\pi G}\right) - kc^2 \qquad (5.27)$$

We have omitted the parenthesized t that should follow R and ρ for the sake of simplicity, but it is still the case that these are time-dependent quantities. Given the values of k and Λ, Equation 5.27 may be used to determine the behaviour of R, although in order to work out the precise details it is also necessary to know the value of ρ at some particular time. In fact, it can be shown that in a pressure-free universe $\rho(t)[R(t)]^3$ has a constant value, D say, so the density information is often provided by specifying the value of this constant, and replacing ρ in Equation 5.27 by D/R^3.

Note that in the case of the critical model, where $k = 0$ and $\Lambda = 0$, the Friedmann equation implies that

$$\dot{R}^2 = \frac{8\pi G R^2}{3}\rho \qquad (5.28)$$

Identifying ρ as ρ_{crit} in this case, and recalling that $[H(t)]^2 = \dot{R}^2/R^2$, the above equation can be rearranged to give Equation 5.26, $\rho_{\mathrm{crit}}(t) = 3H^2(t)/8\pi G$.

It is worth pointing out that on the basis of Newtonian physics (rather than general relativity), the Friedmann equation (Equation 5.27) can be derived by considering the total energy of an expanding spherical distribution of galaxies. In such a derivation kc^2 is related to the total energy of the sphere, $\dot{R}^2$ is related to the kinetic energy of the sphere, and the term involving G is related to the gravitational potential energy of the sphere. On this basis the Friedmann equation is sometimes referred to as the 'energy equation' of the Universe.

$$\Omega_{\mathrm{m}}(t) = \frac{\rho(t)}{\rho_{\mathrm{crit}}(t)} \quad \text{where } \rho_{\mathrm{crit}}(t) = \frac{3H^2(t)}{8\pi G} \qquad (5.29)$$

If, at some time t, the density of matter in the Universe was half the critical value, then $\Omega_{\mathrm{m}}(t) = 0.5$; if the actual density was one-quarter of the critical density then $\Omega_{\mathrm{m}}(t) = 0.25$, and so on. Note that this parameter includes *all* kinds of matter: dark matter as well as luminous matter, baryonic as well as non-baryonic.

Interestingly, it is possible to represent the value of the cosmological constant in a similar way. If you look at the Friedmann equation (Equation 5.27), you will see that the constant $\Lambda c^2/8\pi G$ enters the equation in a similar way to the matter density ρ. This suggests that we can interpret the term $\Lambda c^2/8\pi G$ as a sort of 'density' associated with the cosmological constant. Of course, $\Lambda c^2/8\pi G$ is a very odd sort of density because it is expected to remain constant while the Universe expands, whereas the matter density ρ is expected to decrease in proportion to $1/R^3$ in an expanding Universe. Nonetheless, if we use the symbol ρ_Λ to represent $\Lambda c^2/8\pi G$, then we can define the **density parameter for the cosmological constant** as

$$\Omega_\Lambda(t) = \frac{\rho_\Lambda}{\rho_{\mathrm{crit}}(t)} \quad \text{where } \rho_\Lambda = \Lambda c^2/8\pi G \qquad (5.30)$$

Note that although Λ and ρ_Λ are constants, the density parameter $\Omega_\Lambda(t)$ does depend on time, because it involves the critical density, and that is time-dependent.

Now, even though ρ_Λ doesn't make much sense as a matter density, if we multiply it by c^2 we obtain a quantity that can be measured in units of J m^{-3} (i.e. joule per

cubic metre). This quantity, $\rho_\Lambda c^2 = \Lambda c^4/8\pi G$, can be interpreted as an *energy density*, where the energy concerned can be thought of as a property of space itself – the energy density of a vacuum! As space expands there would be no reduction in the density of this particular kind of energy. Rather, an increase in the volume of space would simply produce a corresponding increase in 'vacuum energy'. There is no real need to interpret the cosmological constant in terms of a vacuum energy, but it is a fascinating thought that the effect of the cosmological constant might actually be caused by a vacuum energy. It has even been proposed that the vacuum energy might be the result of some sort of quantum physical effect in empty space. Attempts to formulate detailed theories along these lines have not been particularly successful, so rather than assuming that the cosmological constant is caused by some particular kind of vacuum energy, many scientists prefer to call the energy that *may* be associated with the cosmological constant **dark energy** and to admit that the nature of this enigmatic energy is still a complete mystery. Even so, it is now common practice to quote values for the dark energy density rather than for the cosmological constant itself. It is also common practice to refer to $\Omega_\Lambda(t)$ as the density parameter for dark energy.

A lot of observational effort is now going into the determination of the current values of $\Omega_\mathrm{m}(t)$ and $\Omega_\Lambda(t)$. Figure 5.30 gives some indication of the significance of these two parameters, which can be denoted $\Omega_{\mathrm{m},0}$ and $\Omega_{\Lambda,0}$. At all points on the red line, $\Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} = 1$. If the observed values of $\Omega_{\mathrm{m},0}$ and $\Omega_{\Lambda,0}$ satisfy this condition, then the geometry of space will be flat, and it must be the case that $k = 0$ (Question 5.9 asks you to prove this). On the other hand, if $\Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} > 1$ then $k = +1$, or if $\Omega_{\mathrm{m},0} + \Omega_{\Lambda,0} < 1$ then $k = -1$. Thus, the geometric properties of space depend crucially upon the sum of $\Omega_{\mathrm{m},0}$ and $\Omega_{\Lambda,0}$. As Figure 5.30 also indicates, another

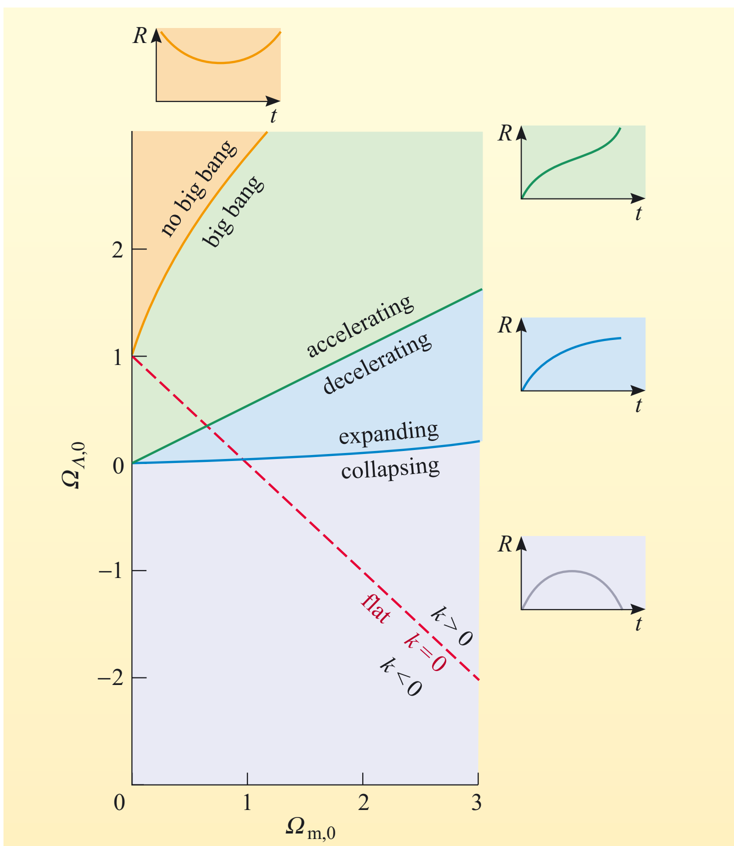


Figure 5.30 A plot of $\Omega_{\Lambda,0}$ against $\Omega_{\mathrm{m},0}$. The values of these two quantities determine important characteristics of a Friedmann–Robertson–Walker cosmological model, such as the sign of the curvature parameter k (red line), whether the Universe will expand forever and eventually collapse (blue line), whether that expansion will accelerate or decelerate (green line), and whether or not there was a big bang (yellow line).

condition involving $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ (represented by the blue line) determines whether the Universe will eventually collapse or continue expanding forever. And, if the fate of the Universe is perpetual expansion, yet another condition involving $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ (represented by the green line) will determine whether the expansion will accelerate or decelerate.

The current values of these density parameters are not easy to determine. However, recent measurements have indicated that $k = 0$, implying that $\Omega_{m,0} + \Omega_{\Lambda,0} = 1$, while other measurements suggest $\Omega_{\Lambda,0} > 0$. In fact, current estimates (see Chapter 7 for details) indicate that $\Omega_{m,0} \approx 0.3$ and $\Omega_{\Lambda,0} \approx 0.7$. If these figures are correct, then most of the energy in the Universe is dark energy, space has a flat geometry, and cosmic expansion will continue forever at an accelerating rate.

QUESTION 5.9

Show that the Friedmann equation can be rewritten as

$$H^2 = \frac{8\pi G}{3}\left(\rho + \frac{\Lambda c^2}{8\pi G}\right) - \frac{kc^2}{R^2}$$

and that this may itself be rewritten as

$$\Omega - 1 = \frac{kc^2}{R^2 H^2} \qquad (\text{where } \Omega = \Omega_m + \Omega_\Lambda)$$

Hence justify the statement that if $\Omega_m + \Omega_\Lambda = 1$, then $k = 0$.

QUESTION 5.10

The Friedmann equation, together with the relation ρR^3 = constant, may be used to show (you are not expected to demonstrate this)

$$2\dot{R}\ddot{R} = -\frac{8\pi G}{3}(R\dot{R}\rho - 2R\dot{R}\rho_\Lambda)$$

Use this, and the definitions given above for $H(t)$, $q(t)$ and $\rho_{crit}(t)$ to show that at any time t

$$q(t) = \frac{\Omega_m(t)}{2} - \Omega_\Lambda(t)$$

Use this, together with the data given above, to estimate the value of q_0.

5.4.4 The Hubble time and the age of the Universe

The critical model not only provides a useful reference value for the cosmic density, it also provides useful insights into the age of the Universe. The usefulness of the critical model stems from the fact that its scale parameter varies with time in a very simple way

$$R(t) = At^{2/3} \qquad (\text{where } A \text{ is a constant})$$

(This relationship is found by solving the Friedmann equation with $k = 0$ and $\Lambda = 0$.)

Combining this with the definition of the Hubble parameter, $H(t)$, it can be shown that in the critical model

$$H(t) = \frac{2}{3t} \qquad (\text{critical universe only})$$

It follows from this that observers living in a universe that was well described by the critical model would find that, after their universe had been expanding for a time t_0, their observations of distant galaxies would indicate that the Hubble constant had the value $H_0 = 2/3t_0$. In other words, the observers in such a hypothetical universe would be able to deduce the age of their universe by measuring H_0 and using the relation

$$t_0 = \frac{2}{3H_0} \quad \text{(critical universe only)}$$

Since H_0 may be expressed in units of s^{-1}, the quantity $1/H_0$ may be expressed in time units, such as seconds or years. The quantity $1/H_0$ is known as the **Hubble time** and is often used as a reference value in discussions of cosmic age, just as ρ_{crit} is a useful reference value in discussions of cosmic density. The exact value of the Hubble time is somewhat uncertain, due to the uncertainties that still exist in measurements of H_0, but it is thought to be about 4.3×10^{17} s or, if you prefer, about 14 billion years (i.e. 1.4×10^{10} yr).

■ If our Universe was well represented by the critical model, how old would it be?

❑ According to the critical model the age of the Universe, t_0, is two-thirds of the Hubble time. Since the Hubble time for our Universe is about 14 billion years, it follows that the age of our Universe, if it were well represented by the critical model, would be about 9 billion years. In fact, this is too short to be realistic.

The critical model is unusual in providing such a simple relationship between the age of the Universe and the observed value of the Hubble constant. Similar relationships exist in other FRW models, but they are generally less direct and therefore more complicated. Rather than trying to write down those relationships it is much easier to represent them graphically. First however, take a look at Figure 5.31, which should give you a general feel for what you can expect to see later.

Figure 5.31 shows the growth of the scale factor for four different FRW models; the models are numbered 1 to 4, and are, respectively,

1 a closed model with $\Lambda = 0$ and $k = +1$

2 a critical model with $\Lambda = 0$ and $k = 0$

3 an open model with $\Lambda = 0$ and $k = -1$

4 an accelerating model with $\Lambda > 0$ and $k = 0$

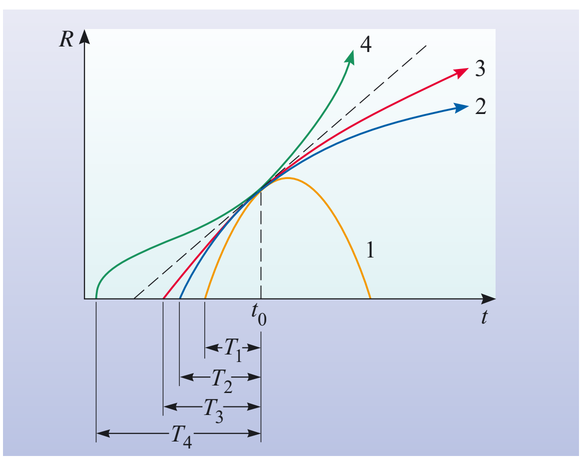


Figure 5.31 The evolution of the scale factor in closed, critical, open and accelerating FRW models that have the same Hubble constant at time t_0. The accelerating model, the only one of the four to have a non-zero cosmological constant, has the greatest age.

The general behaviour of these four models was shown in Figure 5.23, but as redrawn in Figure 5.31 the curves have been shifted horizontally so that they all have the same values for R and $\dot{R}$ at the time t_0 that corresponds to the present. This amounts to saying that the curves have been drawn in such a way that they all indicate the same value for the Hubble constant. Given that the four curves in Figure 5.31 all correspond to the same Hubble constant, what do they tell us about the ages of these four kinds of universe? Well, the age of each model universe is represented by the time that elapses between the moment when R was first equal to zero and the time t_0. These times are different in the four models and are indicated by the values T_1, T_2, T_3 and T_4. As you can see, each is larger than its predecessor, with the closed universe having the smallest age and the accelerating universe the greatest.

You saw earlier that in our Universe the observed value of the Hubble constant is so large that a critical model would imply an unrealistically small age for the Universe. A number of cosmologists have taken this fact alone as *prima facie* evidence that the cosmological constant is non-zero.

Now look at Figure 5.32; this puts the implications of Figure 5.31 into a more general context by plotting curves that show the relationship between the age of the universe (i.e. the length of time it has been expanding) and the Hubble parameter in three different FRW models. The curves represent critical and open models with $\Lambda = 0$, and an accelerating model with the same density as the open model, but with a positive cosmological constant. It can clearly be seen that, for a given value of the Hubble parameter, the accelerating model is always the one that has been expanding longest.

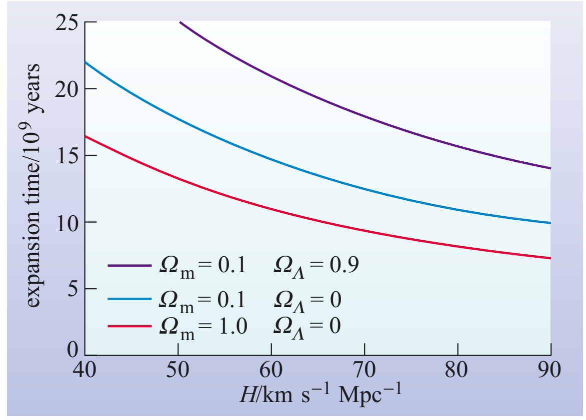


Figure 5.32 The age of the Universe plotted against the Hubble parameter for critical, open and accelerating universes of various densities. (Adapted from Roth, 1997, based on work by J. Primack)

Finally, Figure 5.33 provides an even broader context by showing the age of the universe for any value of the Hubble constant and for a wide range of possible values for $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$. The possible values of the Hubble constant do not appear anywhere in the diagram, but the curved lines that sweep across the diagram indicate the ages of various models measured in multiples of the Hubble time, and this latter quantity implicitly depends on the value of Hubble constant. Note that the axes of Figure 5.33 are $\Omega_{m,0}$ and $\Omega_{m,0} + \Omega_{\Lambda,0}$, so the condition for flat space ($k = 0$) is now represented by the red vertical line through $\Omega_{m,0} + \Omega_{\Lambda,0} = 1$.

It was mentioned earlier that the currently favoured values of the density parameters for matter and dark energy are $\Omega_{m,0} \approx 0.3$ and $\Omega_{\Lambda,0} \approx 0.7$. These values mean that our Universe is represented by the large red dot in Figure 5.33. You can see from the figure that this location implies an age that is slightly less than the Hubble time, which was earlier estimated to be 14 billion years. This age estimate seems to be consistent with the ages of the oldest known objects in the Universe, the globular cluster stars mentioned in Chapter 1.

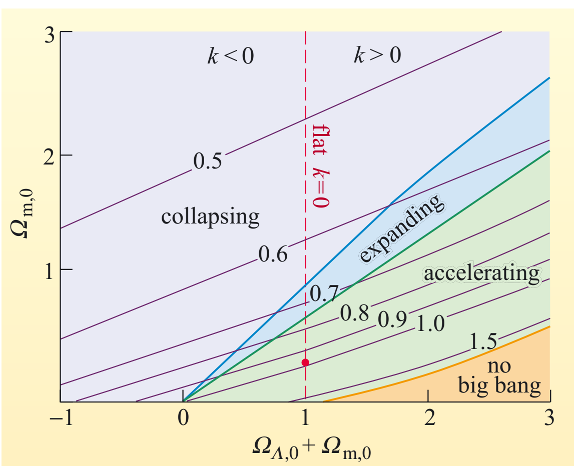


Figure 5.33 The age of the universe in units of the Hubble time is indicated by the various curves that cross this plot of $\Omega_{m,0}$ against $\Omega_{m,0} + \Omega_{\Lambda,0}$. (Adapted from Carroll *et al.*, 1992)

5.5 Summary of Chapter 5

The nature of the Universe

- The Universe contains matter. About 5/6ths of it is believed to be dark matter that may be non-baryonic, and the remaining 1/6th is baryonic matter. The baryonic matter is mainly hydrogen (~75% by mass) and helium (~25% by mass).
- The Universe contains (electromagnetic) radiation. Much of it is visible light, but the major part of the energy is contained in the cosmic microwave background (CMB) radiation.
- The Universe is uniform. That is to say, all regions that are sufficiently large to be representative have the same average density and pressure, wherever they are located. This claim is consistent with the observed distributions of matter and radiation.
- The Universe is expanding. As a result, the redshifts of distant galaxies are found, on average, to be proportional to the distances of those galaxies. This is described by Hubble's law, $z = (H_0/c)d$, where Hubble's constant, H_0, provides a measure of the rate of cosmic expansion at the present time.

Relativistic cosmology and models of the Universe

- According to Einstein's theory of general relativity the geometric properties of space–time are related to the distribution of energy and momentum within that space–time. The precise relationship is described by the field equations of general relativity, which provide the basis for Einstein's theory of gravity and for relativistic cosmology.
- The geometric properties of space–time include curvature. In a curved space, geometric results can take on unfamiliar forms. The interior angles of a triangle may have a sum that is different from 180°, straight lines may close upon themselves, and pairs of straight lines that are initially parallel may converge or diverge. The geometric properties of any particular space–time can be summarized by writing down an appropriate four-dimensional generalization of Pythagoras's theorem. In the case of a static (i.e. non-expanding), flat (i.e. zero curvature) space–time this takes the form

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2 - c^2(\mathrm{d}t)^2$$

- The distribution of energy and momentum throughout space–time is believed to be uniform on the large scale. This assertion is given precise form by the cosmological principle according to which, on sufficiently large size scales, the Universe is homogeneous and isotropic. Simple cosmological models that are consistent with this principle assume that a gas uniformly fills the Universe. Describing the state of this gas involves specifying its density and pressure, $\rho(t)$ and $p(t)$, both of which are expected to change with time due to the expansion or contraction of the Universe.
- In applying general relativity to cosmology, Einstein introduced a cosmological constant Λ. Thanks to this he was able to formulate a relativistic cosmological model that is neither expanding nor contracting, and in which space has a uniform positive curvature. Later, de Sitter presented a model in which the curvature of space was zero, but there was perpetual expansion.

- The work of Friedmann, Robertson and Walker resulted in the specification of the class of cosmological models that are consistent with general relativity and with the cosmological principle. These models involve a curvature parameter k, that characterizes the geometry of space, and a scale factor $R(t)$ that describes the expansion or contraction of space. The full range of FRW models includes cases that are closed, critical, open and accelerating. The Einstein model arises as a special case, and the de Sitter model as a limiting case.
- The behaviour of the scale factor in a pressure-free universe is determined by the Friedmann equation, and depends on the values of k, Λ and the density ρ at some particular time.

Key parameters of the Universe

- The FRW models provide a natural interpretation of the redshifts of distant galaxies as cosmological redshifts caused by the stretching of light waves while they move through an expanding space.
- The Hubble parameter, $H(t)$, provides a measure of the rate of expansion of space in any FRW model. It is defined by $H(t) = \dot{R}/R$, where $\dot{R}$ denotes the rate of change of R. Observations of distant galaxies are predicted to show that, to a first approximation, $d = cz/H_0$, where H_0 represents the value of the Hubble parameter at the time of observation.
- The deceleration parameter, $q(t)$, provides a measure of the rate of decrease of the rate of cosmic expansion in an FRW model. It is defined by $q(t) = -R\ddot{R}/\dot{R}^2$, where $\ddot{R}$ denotes the rate of change of $\dot{R}$. Observations of very distant galaxies are predicted to show systematic deviations from Hubble's law described by

  $$d = (cz/H_0)[1 + (1 - q_0)z/2]$$

  where q_0 represents the value of the deceleration parameter at the time of observation.
- The density parameters Ω_m and Ω_Λ provide a useful means of representing the cosmic matter density and the density associated with the cosmological constant at any time. The parameters are defined by $\Omega_\mathrm{m} = \rho/\rho_\mathrm{crit}$ and $\Omega_\Lambda = \rho_\Lambda/\rho_\mathrm{crit}$ respectively, where ρ is the cosmic matter density at the time of observation, $\rho_\Lambda = \Lambda c^2/8\pi G$ is a 'density' associated with the cosmological constant, and $\rho_\mathrm{crit} = 3H^2(t)/8\pi G$ is the density that the critical universe would have at the time of observation. The quantity $\rho_\Lambda c^2$ can be thought of as the density of dark energy: possibly a 'vacuum energy' associated with empty space. In a Universe with a flat space (i.e. $k = 0$), the Friedmann equation implies that $\Omega_\mathrm{m} + \Omega_\Lambda = 1$ at all times.
- The age of the Universe, t_0, may be conveniently expressed in terms of the Hubble time, $1/H_0$ in any FRW model. In the case of the critical model $t_0 = 2/3H_0$. In other models t_0 may be a different fraction of the Hubble time, depending on the values of Ω_m and Ω_Λ. Increasing the value of Ω_Λ increases the age of the universe for a given value of the Hubble constant.
- The various cosmological parameters are not all independent. The Friedmann equation implies that $\Omega_\mathrm{m} + \Omega_\Lambda - 1 = kc^2/(R^2H^2)$, and it may also be shown that $q = (\Omega_\mathrm{m}/2) - \Omega_\Lambda$.

Questions

QUESTION 5.11

A number of important events in the history of cosmology have been mentioned in this chapter. Compile a chronological listing of these events, starting with the publication of Einstein's theory of general relativity in 1916.

QUESTION 5.12

List the assumptions that underpin the Friedmann–Robertson–Walker models and the Friedmann equation.

QUESTION 5.13

Describe some of the possible consequences of positive curvature in a three-dimensional space, in the context of the FRW models.

QUESTION 5.14

The detailed argument given in Section 5.4.1 showed that the behaviour described by Hubble's law is an expected consequence of expansion in a FRW model. However, in Section 5.4.2 it was stated that this argument was only approximately true because it ignored the acceleration or deceleration of the expansion. Carefully reread the argument in Section 5.4.1 and identify the key step at which acceleration is ignored.

QUESTION 5.15

List the values that have been assigned to all the observational parameters mentioned in Section 5.4. Where a quantity is expressed in more than one unit system, confirm the equivalence of all the given values.